L'IRCOCERVO
PRIMA RIVISTA ELETTRONICA ITALIANA DI METODOLOGIA GIURIDICA - TEORIA GENERALE DEL DIRITTO E DOTTRINA DELLO STATO
FIRST ITALIAN DIGITAL JOURNAL OF LEGAL METHODOLOGY - GENERAL THEORY OF LAW AND DOCTRINE OF THE STATE

Deontología en la decisión penal y nuevas tecnologías

María Luisa de Torres Soto

Universidad Complutense de Madrid

Abstract: **Deontology in Judicial Decisions and New Technologies**
Artificial Intelligence (AI) has been put into practice in all fields of human activity, including the field of Criminal Justice and, therefore, it is already perceived as something accessible. At the same time, the dangers of its inappropriate use are also clear, especially because of the violations that it can cause to the fundamental rights of individuals. To avoid these dangers as far as possible, an ethical and legal approach to AI is needed. One of the valid instruments for the implementation of ethical and legal control of AI is the application of a correct deontology based on prudence, since it can constitute an adequate mechanism for the management of the risks that may arise. This article will attempt to establish the limits of AI in the application of criminal justice, in view of the opportunities and challenges it poses.

Keywords: Justice, Criminal Law, Artificial Intelligence, Risk, Prudence.

Resumen: 1. Introducción – 2. Algunas teorías filosóficas sobre la justicia penal – 3. Referencia a la justicia penal y sistemas algorítmicos: riesgos y desafíos – 4. La prudencia como respuesta deontológica en las decisiones judiciales ante las nuevas tecnologías – 5. Medidas de gestión del riesgo en la aplicación de la tecnología al proceso penal – 6. Algunas ideas a modo de corolario.

1. Introducción

La reflexión sobre la Justicia en el Derecho Penal presenta la centralidad del problema filosófico-jurídico. De hecho, es considerada como "valor superior del ordenamiento jurídico"[1], también regla jurídica para intentar poner "orden social y seguridad en una vida de relación"; otros autores la consideran una cualidad de la conducta humana que expresa su acatamiento a la norma jurídica, es decir, dictada la norma, la conducta del hombre es justa en cuanto se somete a ella e injusta cuando se opone; y, también se afirma que la justicia o la injusticia se produce en función de que se respete "la proporción, el equilibrio, la equidad en la atribución de premios o de sanciones"[2].

[1] Artículo 1 de la Constitución española (CE).
[2] A. Díez Picazo y Gullón, *Sistema de Derecho Civil*, vol. I, 2ª ed., Tecnos, Madrid, 1978.

© L'Ircocervo, 23 (2024) n. 2 ISSN 1722-392X
Special issue: T. Snarski e T.G. Tasso (eds.),
The Concept of Punishment. Justice, Human Rights and Mercy,
https://doi.org/10.57581/LIRCOCERVO.23.2024.2.06

Debido al avance de la técnica, la digitalización y el uso de la IA, considerados actualmente como los principales motores del cambio y la innovación, se esperan importantes avances en la Justicia Penal, en términos de eficacia y eficiencia, pues el derecho sancionatorio no es ajeno a esta nueva realidad y en su modo de aplicar e interpretar la Justicia también se ha dejado asombrar por estas nuevas herramientas tecnológicas.

Ante esta realidad, la prudencia, tan analizada por Aristóteles, debe trasladarse en los albores del siglo XXI sin duda a todo lo relacionado con la tecnología y en especial en el ejercicio que con ella desempeñan los jueces y tribunales de justicia, debiendo prevenir los posibles riesgos, para lo que hay que identificar un *iter* anticipatorio de la prudencia en el uso de las tecnologías por parte de los poderes públicos, incluyendo a los jueces y tribunales de justicia y así justificar deontológicamente su uso, ya que existe el deber moral de abordar dichas tecnologías mediante el previo establecimiento de parámetros éticos y sociales[3] que permitan una aplicación de la IA de carácter universal, que garantice los derechos fundamentales[4].

2. Algunas teorías filosóficas sobre la justicia penal

Sin ninguna pretensión de agotar aquí todas las Teorías y autores que se han ocupado de la Justicia Penal en las distintas Escuelas que se han sucedido en la Filosofía Jurídica, se puede afirmar que la primera consideración filosófica del Derecho y la Justicia, se origina en Grecia, pudiéndose considerar a Aristóteles como el primer filósofo reflexivo sobre el Derecho y el primero que crea una clasificación de la Justicia situándola en su estudio sobre la ética de la virtud[5]. Para el estagirita el orden político y la justicia se orientaban a la consecución de la vida buena en el seno de la comunidad, clasificando la Justicia en *distributiva*, propia de los legisladores, que regula la proporción entre la comunidad de los individuos con respecto a la repartición por aquella de las cargas y beneficios; y *sinalgmática*, propia de los jueces[6], que regula y ordena las relaciones entre iguales, estableciendo

[3] M. Lenca (2021), "Neuroderechos: ¿por qué debemos actuar antes de que sea demasiado tarde?", CIDOP. Recuperado de https://www.cidob.org/publicaciones/neuroderechos-por-que-debemos-actuar-antes-de-que-sea-demasiado-tarde [Fecha de consulta: 30/09/2024], en B.I. Murrugarra Retamozo. *Ivi*, p. 16. "Es esencial que tal prevención se dé por medio de una buena gobernanza de las tecnologías, que abarque una coordinación acertada que cuente con el debido sustento, para lo cual se tiene que recurrir a fuentes confiables".

[4] Incluyendo el derecho a la libertad de pensamiento (cognitiva); a la privacidad de la mente; a la integridad de la mente; a la continuidad psicológica; a la igualdad de acceso a la mejora de la mente; y, a la protección frente al sesgo de los algoritmos. En R. González Álvarez (2021), "'Neuroderechos', prueba neurocientífica y garantía de independencia judicial", in *Derecho & Sociedad* 57 (2021), pp. 1-26 (https://doi.org/10.18800/dys.202102.007).

[5] Aristóteles, *Ética a Nicómaco*, 1929 b), Alianza Editorial, Madrid, 2004.

[6] *Ibidem*.

el equilibrio entre lo que se ha de dar y lo que se ha de recibir en las relaciones privadas, y que incluye la justicia penal, que trata de restablecer la igualdad alterada por el delito, castigando las ofensas igualmente de acuerdo con la gravedad de la falta[7]. Así, para las denominadas teorías absolutas, el sentido de la pena radica más en la retribución, entendiendo, por tanto, que la pena es la consecuencia justa y necesaria del delito cometido, como Kant y Hegel las entendieron, pues el merecimiento fundamenta el castigo.

La Justicia retributiva o denominada concepción moral kantiana, se centra en buscar al culpable y castigarlo. Kant defendía una justificación de la pena basada en la retribución como respuesta a un mal pasado, independientemente de las consecuencias que ello pudiera tener y sostenía que la pena debe aplicarse contra el culpable por la sola razón de que ha delinquido, porque nunca un hombre puede ser tomado por instrumento de los designios de otros. La ley penal es un imperativo categórico. Los requisitos para justificar este retribucionismo son a) que el destinatario de la pena sea responsable por el mal que la pena retribuye; y, b) el mal implícito en la pena sea proporcional al mal que ella retribuye y al grado de responsabilidad del agente.

Las teorías relativas, sin embargo, hablan de la función intimidatoria de la pena, como fin preventivo, según Von Feuerback en el s. XIX como coacción psicológica, que defiende que, ante la posibilidad de recibir un castigo, el sujeto evitará su comisión; o como Franz von Listz, que la entiende como forma de intimidar al que ha cometido delitos, para que no vuelva a cometerlos, apartándolo de la vida social en libertad.

El Derecho español actual, siguiendo a Calus Roxin[8] y a su "Teoría dialéctica de la unión" se basa en que la finalidad de la pena tiene, en un primer momento, una función preventiva y una vez aplicada la pena, la reintegración social del penado[9], ponderando ambos procesos. Por tanto, se parte de la idea de retribución, pero con un fin preventivo, para la protección de la libertad individual y del orden social, y sin olvidar la tarea reeducadora y resocializadora del penado[10].

En ambos procesos intervienen en la actualidad los instrumentos que la técnica, la digitalización y la IA ofrecen para la aplicación de la Justicia Penal como se verá más adelante. Hay que afirmar que los Tribunales deben fundamentar la individualización exacta de la pena atendiendo a la mayor o menor gravedad del hecho y a las circunstancias personales del delincuente. Sin embargo, no en pocos casos se observa en la práctica judicial actual que el Juez no motiva, o al menos no

[7] Posición prácticamente compartida por Tomás de Aquino que sobre la Justicia recoge la definición de Ulpiano ("la perpetua y constante voluntad de dar a cada uno lo suyo") señalando que es la más alta de las virtudes, tal como decía Cicerón.

[8] Cfr. C. Roxin, "Cambios en los fines de la teoría de la pena", in *Teoría del delito en la discusión actual*, traducido por Manuel Abanto Vázquez, Grijley, Lima, 2007. pp. 76-85.

[9] C. Roxin, *Derecho Penal, parte general*, tomo *I. Fundamentos. La estructura de la Teoría del Delito*, 1ª ed., Civitas, Madrid, 1997, p. 95.

[10] STC 19/1988, de 15 de febrero, FJ 9; STC 160/2012, de 20 de septiembre, FJ 3; entre otras.

lo hace suficientemente, la individualización de la pena; ofreciendo como argumentación la fórmula legal en atención a la las circunstancias personales del delincuente y a la mayor o menor gravedad del hecho. En este sentido, Besio Hernández ha realizado un análisis sobre la interpretación que la jurisprudencia está realizando de los criterios relativos a la gravedad del hecho y a las circunstancias personales del delincuente. Este autor llega a la conclusión de que "se observa primacía operativa del baremo legal de gravedad del hecho sobre el criterio de circunstancias personales del delincuente"[11] y propone que las circunstancias personales del delincuente se valoren siempre a favor del reo y, por tanto, que no se utilicen "estos parámetros para aumentar la cuantía de la pena, esto es, que se maneje en perjuicio del reo"[12]. En definitiva, para determinar correctamente la pena, los Tribunales deben estar adecuadamente asesorados y tener a su alcance todos los medios posibles que le permitan valorar el hecho y las circunstancias del delincuente. Es obvio que este proceso no puede convertirse en una operación matemática, porque siempre existirá un margen de ejercicio de discrecionalidad para que el juez adapte la pena al caso concreto. En lo que a este artículo se refiere, deberíamos preguntarnos si las nuevas tecnologías permiten llevar a cabo esta función de auxilio judicial de manera que permitan adaptarse a cada caso concreto.

Entre las teorías de justicia y moralidad social se encuentran las teleológicas y las deontológicas. Las primeras son las que hacen prevalecer los bueno sobre lo moralmente correcto, es decir, juzgan las acciones no por sus cualidades intrínsecas sino por cómo contribuyen a alcanzar cierta meta valiosa o a realizar o maximizar cierto estado de cosas intrínsecamente bueno. Las segundas, también denominadas formalistas, al contrario, dan prioridad a lo moralmente correcto sobre lo bueno, juzgando a las acciones por sus cualidades intrínsecas que las hacen objeto de derechos y deberes y sólo tomando en cuenta su función de satisfacer ciertos objetivos valiosos dentro del marco de tales derechos y deberes[13].

La deontología ha influido en construcciones jurídicas fundamentales, como los derechos humanos y están en la base de propuestas de filósofos como John Rawls o Jürgen Habermas y Dworkin, que han renovado la Filosofía del derecho planteando de nuevo en el ámbito académico la pregunta por la justicia como la virtud principal del orden social. Según Dworkin, es fundamental analizar cada caso dentro de su contexto para descubrir la respuesta adecuada, la cual, según él, sí existe. De esta manera, su referencia a los principios, que actúan como guías para proteger los valores del sistema jurídico, resulta motivadora y permite al juez una mayor flexibilidad en su labor, especialmente cuando no cuenta con una norma específica que aplicar.

[11] M. Besio Hernández, *Los criterios legales y jurisprudenciales de individualización de la pena*, Tirant lo Blanch, Valencia, 2011, p. 407.
[12] *Ivi*, p. 423.
[13] C. Nino, "La valoración moral del Derecho", in Carlos Santiago Nino (a cargo de), *Introducción al análisis del Derecho*, Ariel, Barcelona, 1987, pp. 382-436.

El efecto de la Segunda Guerra Mundial, junto con los juicios por violaciones a la dignidad humana, invita a reconsiderar la noción de justicia desde una perspectiva diferente. Así nació la Carta de Derechos Humanos de la ONU, que define unos estándares mínimos de justicia que son exigibles y que han sido acordados por lo que la mayoría considera como valores fundamentales. Aunque parece que el subjetivismo está muy presente y podría dar la impresión de que cada individuo posee una visión distinta de la justicia, en realidad existe un mayor consenso sobre estos mínimos.

En este sentido, en la actualidad, deberíamos preguntarnos si los nuevos instrumentos digitales y tecnológicos respetan los derechos fundamentales y pueden servir de ayuda a los jueces y tribunales de justicia o más bien, responden a fines utilitaristas, orientado al cálculo de los beneficios y perjuicios de las acciones. Tanto los enfoques utilitaristas de la Justicia como los deontológicos comparten que el tema central de la reflexión ética es la justicia, que la moralidad de las acciones se basa en unos principios y reglas universalmente aplicables a cualquier individuo racional, que la moralidad de un individuo estará en función de su capacidad para seguir dichos principios y reglas, y en su disposición y voluntad de cumplirlos.

Tras los desastres de la II Guerra Mundial, merece la pena destacar la recuperación del pensamiento político de Aristóteles con autores como Hannah Arendt, Alasdair MacIntyre, Charles Taylor y Martha Nussbaum, entre otros. Concretamente, Nussbaum, influida por la Teoría de la justicia de Rawls[14] se esfuerza por reconstruir los principios del universalismo moral, pues la justicia no se limita a la distribución de recursos, sino que se relaciona con la manera en que los valoramos[15] y con el tipo de ciudadanía que aspiramos a promover, defendiendo la justicia y la equidad como dos virtudes fundamentales que deben informar a nuestras instituciones, tanto en su creación como en su cometido[16]. Asimismo, esta autora comparte con Rawls que los Estados deben crear normas neutrales que hagan posible la convivencia basada en criterios de justicia transversales, pues la mayor exigencia moral que tenemos es el compromiso con el futuro de nuestros iguales. Para ello, debería tenerse en cuenta al ser humano de manera global, con sus limitaciones, sus errores, su capacidad de cambio. ¿Podría la IA tener en cuenta la voluntad e intencionalidad de las personas?

Por otra parte, la denominada en la actualidad Justicia *restaurativa* trata de recuperar socialmente al penado y reparar el daño causado a la víctima, o dicho de otro modo subsanar el daño causado por un tercero a un individuo particular en miras a que recupere su bienestar. No se podrá hablar de reinserción si el reo rechaza compensar el daño causado, pues la falta de arrepentimiento manifiesta que su

[14] J. Rawls, *Teoría de la Justic*ia, FCE, México, 2006, p. 69.
[15] Para más información vid. I. Pinedo Cantillo y J. Yañez Canal, "De la justicia social a la justicia compasiva: los aportes de Martha Nussbaum a la filosofía política", in *Discusiones Filosóficas*, 22 (2021), pp. 59-91.
[16] M.C. Nussbaum, *Las fronteras de la justicia. Consideraciones sobre la exclusión*, Paidós, Barcelona, 2002, p. 43.

actuación es moral y jurídicamente reprochable y determinará su capacidad de reinserción. Es conocida la idea de que el Derecho tiene la función de promover o facilitar determinadas conductas para obtener un cambio social, pues la faceta exclusivamente coercitiva sólo permite la conservación de la sociedad. Así, el derecho crea valores cuando promueve e incentiva determinadas conductas que se traducirán en un cambio social para que esas conductas lleguen a consolidarse.

La concepción moral restaurativa se centra más en la responsabilidad del sujeto que en su culpa. La pena queda justificada para promover la felicidad general, desanimando a delinquir y reeducando, de forma que en el futuro se cometan menos delitos, lo que supone un beneficio social. La pena quedaría justificada si es un medio eficaz para evitar la concurrencia de ciertos males sociales si no hay otra forma menos perjudicial para evitar estos males y siempre que el perjuicio que supone para el destinatario sea menor que el perjuicio que la sociedad sufriría si la pena no se aplicara.

Naciones Unidas define la Justicia Restaurativa como "una forma de responder al comportamiento delictivo balanceando las necesidades de la comunidad, las víctimas y los delincuentes"[17]. Se alude a los tres protagonistas: víctima, victimario y sociedad. Se añade "cualquier proceso en que las partes afectadas – ofensor, víctima y miembros de la comunidad – participan en conjunto y activamente en la resolución de los asuntos derivados del delito, generalmente con la ayuda de un facilitador". Para esta Organización, la Justicia Restaurativa es una metodología que permite solucionar problemas a través de involucrar a los principales afectados y del ofrecimiento de ayuda desde la comunidad, tanto para la víctima como para el ofensor. Las normas internacionales favorecen la justicia restaurativa y han abierto el camino hacia una nueva política criminal, hacia otras vías de solución de conflictos, para intentar dar respuesta a todas las necesidades de los protagonistas del proceso penal.

Los programas de justicia restaurativa pueden ofrecer a los agresores una oportunidad para asumir la responsabilidad por la ofensa y entender los efectos de la ofensa en la víctima; expresar las emociones, incluso remordimiento, sobre la ofensa; recibir apoyo para reparar daño causado a la víctima o a un mismo y a la familia; corregir actitudes, restituir o reparar; disculparse con las víctimas; restaurar la relación con la víctima en caso de que sea apropiado; y conseguir cerrar una etapa[18].

Pero como se señalaba anteriormente, nuestro sistema judicial penal se fundamenta en el principio de legalidad y de justicia retributiva[19]. Ello se traduce en que los Cuerpos de Seguridad del Estado, el Ministerio Fiscal y los Jueces están

[17] *Manual sobre Programas de Justicia Restaurativa*, ONU, 2006. Recuperado de: https://www.unodc.org/documents/justice-and-prison-reform/Manual_sobre_programas_de_justicia_restaurativa.pdf, pp. 1-5 y 33-36.

[18] Cfr. A. Llano Torres, *Innovación educativa y justicia restaurativa en las Facultades de Derecho y Educación*, Dykinson, Madrid, 2021.

[19] La Constitución española atribuye el ejercicio exclusivo del *ius puniendi* al Estado.

obligados a actuar, conforme a un proceso debido, cuando tienen noticia de la comisión de unos hechos delictivos y no están legitimados para abandonarlo en tanto persistan los presupuestos materiales que lo han provocado, además de estar fundado en la pena y en el castigo al culpable. Poco o escaso margen deja para dar entrada a la justicia restaurativa[20]. Aun así, también nuestra Constitución introdujo el artículo 25.1 que abandona el concepto de la pena puramente retributivo[21].

Algunos autores consideran que "el concepto de justicia restaurativa ha sufrido
intensos cambios, hasta encontrar lo que hoy es el verdadero camino hacia una cultura de superación de contiendas, en lo que se puede llamar teleología de la paz. Conscientes de que esta justificación necesitaría muchas más páginas explicativas, corresponde señalar que la idea de "tribunales de transición", es decir, de intentos de realización de juicios para poner punto final a los conflictos, no son novedad en nuestra historia"[22].

Estos sistemas de justicia no podrían ser sustituidos por la IA pues aunque sí podrían utilizarse para facilitar en algunas funciones predictivas el trabajo de los juristas, la misericordia ante el penado con base en las circunstancias personales del delincuente y su oportunidad de reinserción, pertenecen sólo a la mirada humana y no pueden estar presentes en los sistemas de IA.

Por el contrario, señala Beorlegui entre otros, que el cientifismo reduccionista o determinismo materialista considera que la racionalidad científica y técnica es soberana y que es la única capaz de indicar las verdades epistemológicas de los humanos. Por tanto, considera un error la introducción de estándares filosóficos y éticos pues los imperativos tecnológicos suponen que la ciencia tiene su propio funcionamiento lógico y que cualquier injerencia externa obstaculizaría el progreso de la investigación científica y tecnológica en sus diversas aplicaciones[23]. Y

[20] En España, véase: P. Tapia Ballesteros, M.A. Iglesias Rios, "Aproximación a la Justicia Restaurativa en el ámbito penal en España", in *La ley penal: revista de derecho penal, procesal y penitenciario,* 127 (2017); S. Serrano "La integración del Derecho Internacional Humanitario en la Jurisdicción Especial para la Paz", in *Derechos Humanos Justicia Transicional y Perspectivas Globales*; P.A Ramírez Barbosa y V. G. Rodríguez (orgs), Tirant lo Blanch, Bogotà, 2021, p. 320.

[21] El artículo 25.2 de la Constitución española (CE) señala que "las penas privativas de libertad y las medidas de seguridad estarán orientadas hacia la reeducación y reinserción social y no podrán consistir en trabajos forzados". Así, la reeducación y la reinserción son fines mediatos de la pena, instrumentos para conseguir esa reinserción.

[22] V.G. Rodríguez, "Nuevas tendencias del perdón en Derecho penal: utilitarismo, justicia y concreción de la paz", in *AFD*, XXXVIII (2022), pp. 215-250, p. 235.

[23] Vid. C. Beorlegui, "Implicaciones filosóficas y perspectivas éticas de la inteligencia artificial", in R. Amo Usanos, *Inteligencia artificial y bioétic*, Comillas, Madrid, 2023, p. 16. En contraste con esta posición, las consideraciones filosóficas y éticas proporcionan criterios razonables para cambiar la dirección de la investigación tecnocientífica, así como un proceso para detener la investigación tecnocientífica que sea peligrosa y dañina. M. Weber, *La política como Vocación, en el político y en el científico*, Alianza Editorial, Madrid, 1969, p. 176. "Es, por el contrario, infinitamente conmovedora la acción del hombre maduro, de pocos o de muchos años, eso no importa, que siente realmente y con toda su alma esta responsabilidad por las consecuencias y actúa conforme a una

Bayertz subraya que la filosofía debe abandonar la metafísica dogmática y mezclarse con el resto de la racionalidad (…)[24].

3. Referencia a la justicia penal y sistemas algorítmicos: riesgos y desafíos

La IA está ganando una presencia creciente en el trabajo jurídico, impulsando la automatización de una serie de tareas de carácter cognitivo que, hasta hace menos de una década, estaban reservadas a la inteligencia humana"[25]. Ante lo expuesto hasta ahora, cabe plantear la cuestión del papel que desempeña la tecnología, la digitalización y, en general, la IA en el Derecho penal.

Con independencia de lo que tradicionalmente ha sido en la Justicia penal la búsqueda de Legislación, Jurisprudencia y Doctrina para el análisis de las bases jurídicas que fundamentan las resoluciones judiciales, en los últimos años se ha visto incrementado el uso de nuevas herramientas de IA para dicho cometido que permiten ir más allá de la búsqueda de fuentes jurídicas, facilitando la labor del jurista en sus tareas profesionales.

Una de las facetas en la que la IA está introduciéndose en el ámbito penal es en el apoyo a la investigación policial. Las fuerzas de seguridad están empleando herramientas de IA para tareas como reconocimiento facial, análisis predictivo de delitos o procesamiento de grandes volúmenes de datos[26].

En concreto, los análisis predictivos, con sistemas de IA, ayudan a los abogados en sus cometidos como, por ejemplo, para predecir la probabilidad de reincidencia o peligrosidad de un acusado. Un ejemplo es el sistema COMPAS[27] utilizado en Estados Unidos para decisiones sobre libertad condicional[28]. Pero el

ética de la responsabilidad y que, al llegar un cierto momento, dice: no puedo hacer otra cosa, aquí me detengo". En este sentido M. Villas Olmeda, J. Camacho Ibáñez, *Manual de ética aplicada en inteligencia artificial*, Edi. Anaya, Madrid, 2022, p. 25; C. Beorlegui, "Implicaciones filosóficas y perspectivas éticas de la inteligencia artificial", cit., p. 17.

[24] K. Bayertz, "La moral como construcción. Una autorreflexión sobre la ética aplicada", in A. Cortina, D. García Marzá (en cargo de), *Razón pública y ética aplicadas. Los caminos de la razón práctica en una sociedad pluralista*, Tecnos, Madrid, 2003, pp. 47-69.

[25] J.I. Solar Cayón, "La codificación predictiva: inteligencia artificial en la averiguación procesal de los hechos relevantes", in *Anuario Facultad de Derecho de la Universidad de Alcalá*, 11 (2018), pp. 75-105.

[26] C. Alonso Salgado, "Acerca de la inteligencia artificial en el ámbito penal: especial referencia a la actividad de las fuerzas y cuerpos de seguridad", in *IUS ET SCIENTIA*, 7 (2021), pp. 25-36 (https://dx.doi.org/10.12795/IETSCIENTIA.2021.i01.03).

[27] F. Miró Llinares, "Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots", in *Revista de Derecho Penal y Criminología*, 3 (2018), n. 20, p. 108 ss.

[28] El caso State v. Loomis como ejemplo. En la resolución, se da por buena la utilización de la herramienta con ciertas restricciones: no puede utilizarse en la imposición de la condena ni la decisión sobre su dureza; no puede ser por sí misma el elemento determinante de ninguna decisión; y su uso ha de estar precedido de una advertencia al tribunal sobre sus limitaciones técnicas. En J.I. Solar Cayón, "La codificación predictiva", cit., pp. 75-105.

análisis privado y técnico de las acciones futuras de una persona no puede limitar su derecho a la presunción de inocencia, y no solo debido a que podría afectar su privacidad, lo que ya sería un motivo para oponerse a ello, sino también porque el programa de evaluación de su personalidad y la proyección de su comportamiento podría no ser "objetivo".

Pero, la IA está yendo más allá, pues comienza a estar presente también en la función jurisdiccional, estableciendo mecanismos de administración de justicia para la resolución de conflictos de manera automática[29]. Aunque con reticencias, se están desarrollando sistemas para apoyar a los jueces en la valoración de pruebas y toma de decisiones. Países pioneros en este ámbito han sido Canadá, Reino Unido, China, y ya está anunciado en Francia y Estonia. Por otra parte, Argentina está incorporando los análisis predictivos de propuestas de resolución judicial y Colombia incluso para el auxilio a los tribunales en la toma de determinadas decisiones[30].

Entre los desafíos y preocupaciones que presenta la IA en la Justicia penal está el que las computadoras realicen tareas que anteriormente requerían intervención humana o que se consideraban propias de la especie humana porque requerían alguna forma de inteligencia. Esto es una realidad que exige fijar, desde la ética y el derecho, medidas que protejan las facultades, principios y libertades esenciales de los seres humanos respecto a la aplicación de tecnologías que incidan en la actividad cerebral. Precisamente, la introducción de estas tecnologías nos plantea la necesidad de revisión del concepto de "riesgo" y de cómo la intervención de la IA puede alterar la naturaleza de la función jurisdiccional, si se sustituye por la del ser humano.

Se pueden encontrar sistemas algorítmicos de evaluación de riesgos en la administración de justicia penal estadounidense, con el fin de predecir la probabilidad de reincidencia criminal de un delincuente a través de técnicas de *machine learning* que son capaces de identificar una serie de correlaciones entre determinados factores personales y sociales y el riesgo de comisión de futuros delitos. Ello se consigue a través de tener en cuenta indicadores de circunstancias personales como la edad, el sexo, el nivel de estudios, el contexto familiar, la situación socio-laboral; elementos socio-demográficos como el lugar de residencia, el contexto socioeconómico; el historial judicial del individuo, los posibles delitos previos o el historial de violencia; y factores predictores del riesgo de reincidencia, asignando a cada uno de ellos un peso relativo en función del análisis de los casos pretéritos. Estas herramientas proporcionan a los funcionarios públicos encargados de tomar una decisión cuantificando la probabilidad de reincidencia en la comisión

[29] La gran pantalla se ha hecho eco recientemente de este tipo de actividades como se observa en la película "Justicia artificial" estrenada en España el 13 de septiembre 2024, Dir: Simon Casal.

[30] Cfr. J.I. Solar Cayón, "Inteligencia artificial en la justicia penal: los sistemas algoritmos de evaluación de riesgos", in J.I. Solar Cayón (en cargo de), *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho*, J. I. Universidad de Alcalá, Madrid, 2020, pp. 126-127.

de un delito y determinando su nivel de riesgo: – bajo, medio o moderado y alto riesgo[31].

Así, la automatización ha cobrado un auge especial en la justicia penal estadounidense, tanto en la etapa de la investigación policial, para identificar zonas de mayor peligro, como en la fase judicial, empleando programas para analizar la escena del crimen y las pruebas – balística, huellas, reconocimiento facial, ADN – y obtener evidencias para la determinación de los hechos. También en los sistemas de evaluación del riesgo de reincidencia, para resolver sobre la concesión de libertad provisional o para determinar la sentencia y hasta para la ejecución de la pena en un sistema penitenciario crecientemente privatizado y automatizado, en el cual estos mismos sistemas son ya una herramienta habitual para decidir asuntos como el tratamiento de rehabilitación adecuado a cada recluso y la concesión o no de libertad condicional[32].

Ahora bien, a nadie se le escapa la posibilidad de que el sistema subestime el riesgo y recomiende dejar en libertad a delincuentes peligrosos que vuelven a cometer crímenes violentos y ello sería un escándalo que devaluaría el valor comercial de la herramienta y provocaría pérdidas importantes a la compañía propietaria, de modo que ésta podría introducir mecanismos que eviten este tipo de errores – falsos negativos – aun a costa de incrementar los errores del primer tipo – falsos positivos[33].

De hecho, entre los desafíos y preocupaciones que surgen hay que referirse a que existe el riesgo de que los sistemas de IA reproduzcan o amplíen sesgos existentes en los datos con los que son entrenados, generando posibles decisiones discriminatorias. Es decir, los algoritmos de IA también pueden estar sesgados – cometer errores sistemáticos – pues los algoritmos son entrenados a partir de datos que son producto de decisiones humanas pasadas, y si contienen sesgos, el modelo de IA puede aprenderlos y reproducirlos. De hecho, existen ya numerosas pruebas de que, efectivamente, la inteligencia artificial hereda y amplifica los sesgos humanos[34].

Además, se puede vulnerar el derecho de transparencia por el desconocimiento de los modos con los que la IA llega a las conclusiones para poder

[31] *Ivi*, p. 129.
[32] *Ivi*, p. 144.
[33] *Ivi*, p. 145.
[34] L. Vicente y H. Matute, "Humans inherit artificial intelligence biases", in *Scientific Reports* (2023), https://doi.org/10.1038/s41598-023-42384-8. Investigación realizada por las psicólogas Lucía Vicente y Helena Matute de la Universidad de Deusto, Bilbao. "No solo la IA heredaría sesgos de los humanos, como ya sabíamos, sino que también las personas podríamos heredar sesgos de la IA, Estos resultados muestran que información sesgada proporcionada por una inteligencia artificial puede ejercer un impacto negativo perdurable en las decisiones de las personas. El hallazgo señala la necesidad de más investigación psicológica y multidisciplinar sobre la interacción IA-humanos. También se hace necesaria una regulación basada en evidencia que garantice una IA ética y fiable, que no tenga en cuenta solo los aspectos técnicos, sino también los aspectos psicológicos de la relación entre IA y humanos".

ejercer el derecho a la defensa; pueden existir riesgos por no garantizar la presunción de inocencia o la motivación de las sentencias como parte de los derechos fundamentales; y puede generarse responsabilidad patrimonial o civil por la intervención de sistemas de IA autónomos, cuestión que es de gran importancia para las arcas tanto de los entes públicos como privados.

Se hace complicado contradecir el resultado de un algoritmo que en muchos casos, no es fácil de entender ni por los propios profesionales que lo aplican,[35] por lo que se puede deducir que la IA es imperfecta y además no es clara, haciendo que el derecho a la defensa se convierte en un camino de difícil tránsito.

Ante estos antecedentes, la Unión Europea cuenta con la Carta Ética Europea sobre el uso de la IA en los sistemas judiciales y su entorno[36] para el procesamiento automatizado de decisiones y datos judiciales, basados en técnicas de IA y que pueden guiar a los legisladores y profesionales de la justicia cuando se enfrentan al rápido desarrollo de la IA en los procesos judiciales nacionales.

Así, para que la IA siga estando al servicio del interés general, la Comisión Europea para la Eficiencia de la Justicia (CEPEJ) ha identificado en dicho texto los siguientes principios fundamentales que deben respetarse en el campo de la IA y la justicia**:** En primer lugar, el respeto a los derechos humanos, pues debe garantizarse que estas aplicaciones no reproduzcan o agraven esta discriminación y no conduzcan a análisis o prácticas deterministas. En segundo lugar, el principio de calidad y seguridad, pues debe ser posible procesar los datos mediante el aprendizaje automático sobre la base de originales certificados. En tercer lugar, el principio de transparencia en las metodologías y técnicas utilizadas en el procesamiento de las decisiones judiciales mediante la accesibilidad y comprensión de las técnicas de procesamiento de datos y la posibilidad de que las autoridades o expertos independientes realicen auditorías externas. Y, en cuarto lugar, el principio "bajo el control del usuario" para convertir al usuario en un agente activo y sentirse a cargo de sus elecciones.

En particular, el juez debe poder volver en cualquier momento a las decisiones judiciales y los datos que se han utilizado para producir un resultado y seguir teniendo la posibilidad de apartarse de él, teniendo en cuenta las especificidades del caso en cuestión. Cada usuario debe ser informado, en un

[35] *Ibidem*. (…) se muestra de manera empírica los fallos valorativos a los que la inteligencia artificial nos conduce.

[36] Directiva (UE) 2016/680 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por parte de las autoridades competentes para fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, que establece la prohibición de "las decisiones basadas únicamente en un tratamiento automatizado, incluida la elaboración de perfiles, que produzcan efectos jurídicos negativos para el interesado o le afecten significativamente, salvo que estén autorizadas por el Derecho de la Unión o del Estado miembro a la que esté sujeto el responsable del tratamiento y que establezca medidas adecuadas para salvaguardar los derechos y libertades del interesado, al menos el derecho a obtener la intervención humana por parte del responsable del tratamiento" (art. 11).

lenguaje claro y comprensible, de la naturaleza vinculante o no vinculante de las soluciones propuestas por los instrumentos de IA, las diversas opciones posibles y su derecho a la asesoría legal y al recurso ante un tribunal. En definitiva, debe prohibirse el uso de IA como único elemento determinante en decisiones judiciales; garantizarse la supervisión humana significativa; asegurar la transparencia y explicabilidad de los sistemas; y, establecer mecanismos de auditoría y control.

Por su parte, la Resolución del Parlamento Europeo, de 6 de octubre de 2021, sobre la inteligencia artificial en el Derecho penal y su utilización por las autoridades policiales y judiciales en asuntos penales señala que:

> […] la IA ha experimentado un gran avance en los últimos años, convirtiéndola en una de las tecnologías estratégicas del siglo XXI, que puede generar considerables beneficios en términos de eficiencia, precisión y adecuación y que, por consiguiente, aporta una transformación positiva a la economía y la sociedad europeas, pero también enormes riesgos para los derechos fundamentales y las democracias basadas en el Estado de Derecho; que la IA no debe considerarse un fin en sí misma, sino un instrumento al servicio de las personas, con el objetivo primordial de aumentar el bienestar humano, las capacidades humanas y la seguridad (Considerando A)

Ya se utiliza la IA para informar las decisiones judiciales relativas a la adopción de medidas cautelares, como la determinación de la prisión preventiva o la libertad provisional del acusado, la fijación de la fianza y la imposición de condiciones más o menos rigurosas de control de aquella libertad. Es decir, se trata de informar la decisión judicial, no de sustituir el criterio del juez, sino de proporcionarle información adicional para que pueda adoptar las medidas cautelares más oportunas. Su empleo se centra por tanto en predecir el riesgo de incomparecencia ante el tribunal, el riesgo de reincidencia criminal y, específicamente, el riesgo de comisión de un delito violento[37]. Se sustituye con estos sistemas la noción de "peligrosidad criminal", como característica psicológica – individual y subjetiva – del delincuente, como elemento esencial de valoración del riesgo, por un enfoque actuarial que se centra en el "riesgo de delinquir" que presenta la conducta de una persona de forma objetiva[38]. Metodologías que son diferentes, pero no incompatibles.

En definitiva, se entiende que estas herramientas tienen un carácter preventivo, basados en una forma de "justicia actuarial" cuyo objetivo es del de informar la toma de decisiones que implican ciertos márgenes de discrecionalidad en la aplicación del Derecho para evitar futuras infracciones legales, para lo que utilizan conceptos y métodos análogos a los de la matemática actuarial para la evaluación de riesgos. De este modo, la justicia actuarial presupone un modelo

[37] J.I. Solar Cayón, "Inteligencia artificial en la justicia penal…", cit., p. 130.
[38] C.M. Romeo Casabona, "Riesgo, procedimientos actuariales basados en inteligencia artificial y medidas de seguridad", in *Revista de Derecho, Empresa y Sociedad*, 13 (2018), pp. 40-41.

conceptual de sistema penal centrado fundamentalmente en la gestión de riesgos futuros, antes que en la modificación de las conductas delictivas. Estos sistemas no pueden ser utilizados por sí solos para agravar o atenuar la condena, pero sí pueden ser tomados en consideración para alcanzar una sentencia informada y ajustada a las circunstancias de cada condenado: para la suspensión total o parcial de la condena, para el diseño de un programa individualizado para lograr en el futuro la libertad condicional, para la derivación a programas de tratamiento alternativos a la prisión, entre otros fines[39].

Sin embargo, algunos Estados han permitido que los resultados de la evaluación de riesgo proporcionada por estos sistemas constituyan un elemento de juicio más para la determinación y graduación de la condena[40]. Si se emplea un sistema de este tipo con alcance general en el proceso penal y, en los casos en que el acusado sea evaluado como de "bajo riesgo" o "alto riesgo", la Comisión recomienda al tribunal la apertura de una investigación para obtener información adicional sobre el carácter y las circunstancias del acusado que guíe al tribunal en la determinación de la sentencia adecuada en cada caso, sin que los resultados del programa agraven o atenúen la condena, sino sólo para que se abra una investigación con el fin de obtener información adicional sobre el individuo, de forma que se evalúe toda la información en su conjunto[41].

La finalidad de estos programas es la reducción de costes en Áreas en la que se considera que los sistemas de evaluación de riesgos puedan jugar un papel importante para diseñar políticas públicas más "inteligentes" que permitan reducir la población reclusa sin que ello repercuta negativamente en la seguridad pública.

Para poder afinar más en la exactitud de los datos obtenidos mediante la IA, el Libro Blanco sobre la IA de la Comisión Europea, establece que en relación a aquellas aplicaciones de IA que entrañen un riesgo elevado como serían las empleadas en el ámbito judicial el marco regulador podrá exigir, entre otros requisitos, la conservación de registros exactos sobre el conjunto de datos utilizados para entrenar el algoritmo sus características, metodologías de programación, nivel de exactitud esperado, etc…[42].

Junto a las exigencias de transparencia en el proceso de diseño e implementación de los sistemas algorítmicos, otras estrategias de control pasan por la definición de procedimientos y métricas para la monitorización y supervisión de

[39] Como refleja la decisión del Tribunal Supremo de Indiana en Malenchik v. State (2010). Malenchik v. State, nº 79S02-0908-CR-365 (Indiana Supreme Court, 2010), pp. 10-11, cit. en J.I. Solar Cayón, "Inteligencia artificial en la justicia penal …", cit., p. 132.

[40] J.I Solar Cayón, "Inteligencia artificial en la justicia penal …", cit., p. 133: Virginia, en delincuentes sexuales y tráfico de drogas, fraude y hurto sin violencia; Florida, para menores de edad; Oklahoma, para la selección de condenados a los que es posible sustituir la pena privativa de libertad por otra alternativa.

[41] *Ivi*, p. 134.

[42] Cfr. COMISIÓN EUROPEA, Libro blanco sobre la inteligencia artificial: un enfoque europeo orientado a la excelencia y la confianza, COM (2020) 65 final, Bruselas, 19 de febrero de 2020, pp. 23-25 y 6.

sus *outputs*, fundamentalmente al objeto de detectar y corregir posibles sesgos que afecten a la imparcialidad de sus resultados[43]. A la vista de estas consideraciones, resulta patente que el control o la validación de los resultados de los sistemas de *machine learning* es esencial para identificar y mitigar aquellos sesgos originados como consecuencia del propio proceso de aprendizaje del algoritmo y que pueden resultar difíciles o imposibles de detectar en la fase de diseño[44].

Si bien parece lógico defender que los sistemas de algoritmos no han de tomar en consideración atributos constitucionalmente protegidos, como el género o la raza, lo cierto es que en ocasiones puede resultar necesario hacerlo para asegurar la justicia de los resultados. Si diversos grupos raciales o mujeres y hombres tienen diferentes tasas base de reincidencia criminal, la inclusión de estas diferencias podría resultar necesaria para lograr unos resultados más correctos o ajustados a la realidad[45].

En relación a la utilización de este tipo de sistemas es preciso distinguir entre el problema estadístico de la correcta evaluación de riesgos y la cuestión de la política pública acerca del curso de acción apropiado, puesto que aquellos proporcionan una estimación de riesgos pero no determinan el tipo de intervención a llevar a cabo, ni tan siguiera que el tipo de intervención haya de ser la misma en todos los casos con idéntica evaluación de riesgo. Pero siempre imperará el principio de libre apreciación de la prueba por parte del juez[46], debiendo establecer una serie de cautelas para que los jueces sean conscientes de sus potenciales incorrecciones y sesgos[47].

4. La prudencia como respuesta deontológica en las decisiones judiciales ante las nuevas tecnologías

Al pensar en la ética de la IA deben recordarse las leyes de la robótica que imaginó el escritor Isaac Asimov[48]:

> 1º- Un robot no hará daño a un ser humano ni permitirá que, por inacción, esta sufra ibiddaño.
> 2º- Un robot obedecerá las órdenes que reciba de un ser humano, a no ser que las órdenes entren en conflicto con la primera ley.
> 3º- Un robot protegerá su propia existencia en la medida en que dicha protección no entre en conflicto con las leyes primera y segunda

[43] J.I. Sola Cayón, "Inteligencia artificial en la justicia penal ...", cit., p. 156.
[44] *Ivi*, p. 159.
[45] *Ivi*, p. 163.
[46] C.M. Romeo Casabona, "Riesgo, procedimientos actuariales basados en inteligencia artificial y medidas de seguridad", cit., p. 51.
[47] J.I. Sola Cayón, "Inteligencia artificial en la justicia penal ...", cit., p. 172.
[48] I. Asimov, *Circulo vicioso*, Street&Smith, New York, 1942.

Así pues, el planteamiento ético de la cuestión, exige que la tecnología esté al servicio de las personas[49], por eso es importante mantener a éstas siempre en el centro de cualquier análisis y abogar por un marco ético que tenga en cuenta la seguridad y la salud humana, la libertad, la privacidad, la integridad y la dignidad, la autodeterminación y la no discriminación, y, como consecuencia, la protección en cualquier fase del proceso penal.

Es decir, cuando pensamos en la IA y las nuevas tecnologías en general, no basta con centrarnos sólo en las condiciones técnicas de su posibilidad, sino también en sus condiciones éticas, sociales y políticas[50] . Si bien la mayoría de los defensores de la IA reducen la mente humana a una inteligencia lógica[51], no puede olvidarse que, aunque sea posible construir máquinas que actúen como si fueran inteligentes, autónomas y libres, en realidad éstas no son humanas, sino que solamente pueden contener una apariencia de humanidad.

Para Aristóteles la prudencia consiste en ser capaz de poner en práctica el conocimiento teórico del bien. Requiere una voluntad con dominio de lo que se hace, experiencia, conocimiento de los principios morales, reflexión ponderada, atención a las circunstancias, petición de consejo, decisión y ejecución[52]. De ahí, que la prudencia sea necesaria para la deliberación correcta y para el entendimiento y el discernimiento entre lo bueno y lo malo[53]. Así el estagirita, en su obra *Ética a Nicómaco* profundiza en la naturaleza de la misma, considerando a la prudencia como "cualidad que, guiada por la verdad y la razón, determina la conducta que puede ser buena para el hombre, pues éste es capaz de deliberar de manera correcta y conveniente"[54]. La inteligencia se aplica, igual que la prudencia, a problemas que exigen deliberación y conlleva la propuesta de soluciones, pero la prudencia va más allá, pues además de ser normativa, es deliberativa. Es decir, ordena hacer o no hacer algo de manera recta.

En la Filosofía clásica se proponen ocho partes integrales de la prudencia, elementos que necesariamente deben concurrir para el acto perfecto de la misma. Estos elementos son: memoria, entendimiento o inteligencia, docilidad, sagacidad,

[49] Siguiendo a I. Kant, *Fundamentación de la metafísica de las costumbres* (1785) *y Crítica de la razón práctica* (1788): "Obra de tal modo que trates a la humanidad, tanto en tu persona como en la persona de cualquier otro, siempre al mismo tiempo como fin y nunca simplemente como medio".
[50] M. Carabantes López, *Inteligencia artificial: una perspectiva filosófica*, Escolar y Mayo Editores, Madrid, 2016.
[51] C. Beorlegui, *Antropología filosófica. Dimensiones en la realidad humana*, UPCO/Universidad de Deusto, Madrid/Bilbao, 2016. "En realidad, se trata de una realidad mucho más compleja, que involucra no sólo el intelecto sino también los sentimientos y emociones, la autonomía y la libertad, que dan sentido a la existencia".
[52]Aristóteles, *Ética a Nicómaco*, Alianza Editorial, Madrid, 2004.
[53] *Ibidem*. Para ello, Aristóteles señala también la necesidad de conocer lo que es bueno y lo que no lo es, pues la prudencia "no es una ciencia, ni un arte, ni una técnica, es una disposición racional verdadera y práctica respecto de lo que es bueno y malo para el hombre". "Parece propio del hombre prudente discurrir bien sobre lo que es bueno y conveniente. La verdad no necesita cambiar, pero la prudencia cambia constantemente, pues se refiere a lo conveniente en cada caso y para cada uno".
[54] *Ibidem*. Así decía el filósofo "[…] sólo aquel que de antemano es prudente puede ejecutar el bien".

razón, providencia o previsión, circunspección y precaución[55]. El filósofo los divide en dos grupos: 1. Los que pertenecen a la "prudencia cognoscitiva", compuesta por la memoria, entendimiento o inteligencia, docilidad, solercia o sagacidad y razón. 2. Los que pertenecen a la "prudencia preceptiva" – aplicada a la obra –, compuesta por la providencia o previsión, circunspección y la precaución. En concreto, la previsión, ordena algo adecuado al fin*;* la circunspección, tiene en cuenta los distintos aspectos de la situación*;* y, la precaución, evita los obstáculos[56]. Así, para que la razón mande rectamente, debe contar con tres potencias. En primer lugar, sin duda, que ordene algo conveniente al fin (previsión o providencia). En segundo lugar, que atienda las circunstancias del negocio (circunspección). En tercer lugar, que evite los impedimentos (precaución)[57].

Cortina señala que en el marco de una ética cívica de la responsabilidad es en el que cobra sentido el *ethos* de la cautela[58]. Esta actitud debe ser exigida a todos, especialmente a los poderes públicos. En la práctica, distinguir el bien del mal por parte del legislador y el gobernante aunque no cuenten con el conocimiento empírico del científico, podrán hacerlo con los medios jurídicos para no cometer errores o, de no ser posible, al menos, no cometer los innecesarios.

De ahí que se haya afirmado que "la prudencia pertenece a la persona que, además de gobernarse a sí misma, tiene la función de gobernar a la comunidad"[59]. Así la prudencia debe contemplarse no sólo como "valor" sino como "principio". Ambos deben unirse para entablar un concepto que englobe todos los ámbitos y llegar así a la excelencia en las acciones. Ésta podría considerarse la aplicación práctica de la ética política. Por todo ello, es indudable la necesidad de prudencia para inspirar y conducir la labor jurídica de los poderes públicos[60].

Entre esos nuevos principios destacan: (i) Responsabilidad, que se aplica a las empresas, a las que fabrican artefactos de IA y a las que los usan y, tanto a nivel individual como social; (ii) Privacidad, ante la importancia que en la actualidad está tomando el manejo de los *big data* (datos masivos) ya que se impone el respeto a la privacidad, como derecho básico de la Declaración Universal de Derechos Humanos (DUDH) como garantía de la dignidad del individuo; (iii) Equidad, pues en la medida en que se está produciendo el fenómeno del *machine learning* en el manejo de los *big data* (máquinas que, utilizando algoritmos aprenden con el manejo de los datos, siendo capaces después de hacer generalizaciones, una vez realizado el aprendizaje, se pueden dar procesos de discriminación y desigualdad:

[55] *Ibidem* (II-II 48 q).
[56] J.F. Sellés, "La virtud de la prudencia según Tomás de Aquino (I)", in *Cuadernos de Anuario Filosófico*, Servicio de Publicaciones de la Universidad de Navarra, 90 (1999), haciendo referencia a cfr. *Summa Theologiae*, II-II ps., q. 48, a. 1, co y *In Metaphysicorum*, l. 1, cap. 1, n. 11.
[57] Tomás de Aquino, *Summa Theologiae*, II-II 48.
[58] A. Cortina "Fundamentos filosóficos del Principio de Precaución", in C.M. Romeo Casabona (a cargo de), *Principio de Precaución, Biotecnología y Derecho*, Comares, Granada, 2004, p. 16.
[59] Tomás de Aquino, *Summa Theologiae* (II-II, 50-1).
[60] L. Avitabile "I principi del diritto. Una riflessione con Francesco Carnelutti", in *L'Ircocervo*, 22 (2023), n. 1, pp. 41-49.

aparición de sesgos no previstos ni deseados; (iv) Transparencia, como base de la confianza en los poderes públicos. La transparencia también supone el compromiso de quienes promueven productos o actividades potencialmente de riesgo para difundir los estudios que hayan efectuado acerca de la magnitud de los riesgos potenciales y de los esfuerzos que han hecho con vistas a reducir al mínimo o a eliminar tales riesgos; (v) Explicabilidad, que es un principio clave para la IA confiable. Los procesos de IA deben ser capaces de dar las explicaciones necesarias para que las personas afectadas por esas decisiones confíen en el proceso automático. Esto es así porque cada vez hay más situaciones en que los artefactos pueden tomar decisiones más autónomas al margen de la decisión de los humanos y hay que estar seguros de que se respetan nuestros derechos[61].

5. Medidas de gestión del riesgo en la aplicación de la tecnología al proceso penal

Las medidas de gestión del riesgo se adoptan ante el surgimiento de daños que tengan un considerable peligro real o potencial[62]. En la IA nos encontramos actualmente con riesgos inciertos pues su desarrollo posibilita acciones que se escapan del control en muchos casos de la intervención humana. Además, se añaden costes jurídicos, en lo que respecta a la vulneración de valores y derechos fundamentales como la dignidad, el libre desarrollo de la personalidad, la igualdad, la libertad, el honor, la intimidad, la propia imagen y la privacidad, entre otros. Todo ello puede adquirir mayor gravedad si el daño causado resultara irreversible.

Surge por tanto un nuevo modo, radicalmente democrático de decisión política y al mismo tiempo una visión más global de nuestra responsabilidad. Como afirma Jonás, "si antes sólo éramos responsables respecto de las personas que nos rodeaban y en el marco del corto plazo, ahora tenemos una nueva responsabilidad, que tiene por objeto no sólo a los individuos actualmente existentes, sino también a las generaciones futuras y a la humanidad en su conjunto". Jonás concibe la responsabilidad como un principio de gestión absolutamente previo al riesgo. Para él hay dos objetivos: por una parte, procurar que la humanidad siga existiendo en el futuro, y por otra, salvaguardar la calidad propiamente humana de esta existencia en los siglos venideros"[63]. Este autor plantea la necesidad de preguntarse antes de sacar al mercado una técnica determinada, si no podría desembocar en un futuro a largo, incluso a muy largo plazo, en un cuestionamiento de la existencia de la especie humana, o al menos, en una alteración profunda de la calidad de esa

[61] *Ivi*, pp. 23-24.
[62] A. Cortina, "Ética de las biotecnologías", in Isegoría, 27 (2002), pp. 73-89, https://doi.org/10.3989/isegoria.2002.i27.555.
[63] H. Jonás, *El principio de responsabilidad. Ensayo de una ética para la civilización tecnológica*, Herder, Barcelona, 1995. Y.H. Jonás, *Técnica, medicina y ética. La práctica del principio de responsabilidad*, Edit. Paidós, Barcelona, 1996.

existencia. Ante tal desconocimiento, indica que es necesario ponerse en las peores consecuencias posibles y si imaginamos que de ese descubrimiento técnico deriva una alteración profunda de la calidad de la vida humana e incluso de la desaparición de la humanidad, más vale renunciar a esa técnica a pesar de los beneficios que pudieran resultar de ella en lo inmediato, y, en consecuencia, abstenerse.

Por su parte, Beck define el riesgo como una amenaza que puede devenir real, que él denomina "sociedad del riesgo" y que concreta en una serie de principios que a efectos de este estudio cabe considerar también aplicables a la IA y entre los que destacan especialmente los siguientes:

> 1) […] El riesgo es una amenaza, una realidad virtual que puede devenir real. La posibilidad de tornarse una realidad supone que haya una decisión y, consecuentemente, una responsabilidad.
> […]
> 4) Riesgos y percepción de riesgos son consecuencias de la lógica del control. El futuro cada vez más escapa al control humano, por más que se intente colonizarlo. Una serie de cálculos, a través de estadísticas, escenarios, probabilidades de accidentes, cálculos actuariales y estándares u organizaciones de medidas, pueden ser utilizados para intentar tornar predecible lo impredecible e intentar controlar el futuro.
> 5) Los riesgos son materializados por un conflicto entre el conocimiento y el desconocimiento, pues los riesgos pueden surgir de ambos. El avance de la genética, por ejemplo, permite descubrir enfermedades que uno, todavía sano, no podría imaginar que fuera portador, adelantando el conocimiento y la percepción del riesgo.
> […]
> 7) El riesgo puede tornarse invisible. Hay un espacio temporal (también denominado jaula temporal) entre el impacto que produce el riesgo y la manifestación de los efectos no deseados por él causados. El impacto no está directamente vinculado a su verdadero origen pues éste puede ser lejano y hacer difícil el establecimiento del nexo. Esto hace que nadie sepa o quiera saber el origen del riesgo y el no reconocimiento del mismo favorece el surgimiento de efectos no deseados.
> 8) El hombre ha creado un mundo artificial (híbrido) que ya no permite distinguir lo que es naturaleza y lo que es cultura, es decir, lo que es natural y lo que ha sufrido transformación por influencia del hombre[64].

Beck, observó así, que nos encontrábamos en una nueva etapa, que denominó "modernidad reflexiva" por el cambio que se había producido en la percepción de los riesgos[65].

Así, el nivel de perceptibilidad del daño por la sociedad va a generar lo que se denomina riesgo aceptable o residual y eso es algo que hay que definir en cada caso concreto. Esto no supone eliminar cualquier duda acerca de la peligrosidad de

[64] U. Beck, *La sociedad del riesgo global*, 2ª ed., Ed. Siglo XXI, Madrid, 2006, pp. 20 ss.
[65] U. Beck, *La sociedad del riesgo. Hacia una nueva modernidad*, Paidós, Barcelona, 1998, p. 204.

la persona o de la actividad en cuestión y probar un riesgo cero. Esto es lo que puede ocurrir con la IA en las predicciones y decisiones judiciales[66]. Sería contradictorio exigir la prueba científica de que no existe ningún riesgo en un campo donde predomina la incertumbre científica. Ello supondría reclamar una prueba imposible, una prueba negativa, *probatio diabólica*, pues, eso es prácticamente imposible. De lo que en realidad se trata, es de establecer el grado de probabilidad del daño, así como su magnitud[67], sin llegar a invertir, con carácter general, la carga de la prueba. Es decir, no se puede invertir la carga de la prueba porque no se puede pedir por ley una prueba imposible al carecer de datos suficientes o datos generados por una máquina programada con determinada información y no con elementos de intencionalidad, sensibles y cognitivos propios exclusivamente del ser humano.

El progreso tecnológico es bueno cuando puede beneficiar a la mayoría, pero siempre que no perjudique a la minoría. Así, Kaufmann insiste en que no puede haber una libertad absoluta de la ciencia y la investigación, pues "ésta tiene sus límites inmanentes en los valores del cuerpo, la vida, la autodeterminación y dignidad del hombre, y esas limitaciones subyacen a la aplicación tecnológica, pues ya existían en la investigación básica"[68], por tanto perfectamente aplicables a las potencialidades que puede generar la IA y que sin embargo, escapan a nuestro alcance.

Debido al potencial de esta tecnología y, como consecuencia, a los riesgos asociados a su uso, debe limitarse su campo de actuación mediante una regulación adecuada, aunque sin pretender que ésta agote todos los supuestos, debido a que es un campo en continua transformación, y los cambios no son permanentes ni tienen periodos de tiempo por lo que nos obliga, como señala Gómez-Riesgo a responder a los problemas según se van planteando[69].

Ya decía Tomás Moro que "del príncipe, como de un inagotable manantial, viene a los pueblos la inundación de todo lo bueno y de todo lo malo"[70]. Gobernar es difícil, pues los políticos deben gestionar los riesgos con la mayor prudencia que sea posible, conociendo la percepción social del riesgo en la población y sin

[66] En las llamadas *smart prisons*, la IA se emplea en el diseño y administración de la prisión, así como en la supervisión del día a día de los detenidos (detectando, por ejemplo, conductas ilícitas, actos violentos y posibles fugas). Ver: S. Barona Vilar, *Algoritmización del Derecho y de la Justicia: de la Inteligencia Artificial a la Smart Justice*, Tirant Editorial, Valencia, 2021, p. 683 ss.

[67] R. Andorno, "Validez del principio de precaución como instrumento jurídico para la prevención y la gestión de riesgos", in C.M. Romeo Casabona (a cargo de), *op. cit.*, pp. 17-33.

[68] *Ibidem*.

[69] J. Gómez-Riesco, "Los robots y la responsabilidad civil extracontractual", in B. Andrés, *Derecho de los robots*, Ley Wolters Kluwer, Madrid, 2018, pp. 107-130 y 203-228. "Todo ello sin perjuicio de la contingencia de tales normas y de la dificultad de anticiparse y poder regular con la suficiente seguridad jurídica una realidad en permanente desarrollo y evolución, y que todavía no conocemos en su completa dimensión, por lo que habrá que ir dando respuesta a cada cuestión y problema a medida que esto se vayan planteando".

[70] T. Moro, *Utopía*, traducción: Emilio García Estebanez, Clásicos del pensamiento, Tecnos, Madrid, 2017.

alarmismos innecesarios. Conjugar todo ello implica, sin duda alguna, la asunción de algún tipo de responsabilidad.

Tanto si los cambios legislativos se producen, como ocurre por ejemplo con los neuroderechos para proteger la privacidad mental, la dignidad humana, la identidad, la autodeterminación individual, la soberanía y libertad en la toma de decisiones, como se refleja en la Carta de Derechos digitales, como si no, debe darse una respuesta precautoria con los elementos jurídicos con los que ya se cuenta y, en cualquier caso, la responsabilidad principal no debe olvidarse que recae sobre la participación humana en su uso y no tanto en los fallos de la tecnología.

Los sistemas de IA no son seres con derechos y deberes, no poseen libertad de pensamiento o de decisión. Carecen de intencionalidad y voluntad. Cualquier tipo de responsabilidad civil, administrativa o penal recaerá en quien maneje esos sistemas. Por tanto, es su deber diseñar o entrenar modelos de IA respetando los estándares éticos fijados en una sociedad. En definitiva, la ética les corresponde a los humanos, no a las máquinas.

6. Algunas ideas a modo de corolario

Ni todo lo técnicamente posible debe ser regulado de manera discrecional, ni tampoco todo lo técnicamente posible debe ser regulado de forma restrictiva, precisamente, porque están en juego los derechos fundamentales de las personas.

El debate se encuentra por tanto en valorar si se trata de remediar carencias de los sistemas actuales de aplicación de la justicia o si lo que se pretende es alcanzar capacidades más propias de un ser superdotado por una máquina y tomar decisiones que sustituyan la función de los jueces y ahorrar costes en los puestos de trabajo o contar con el riesgo de establecer una ética y un derecho en la tecnología sin una visión general del ser humano, o en sentido más técnico, sin una axiología que la justifique.

Aunque la IA ofrece herramientas potencialmente valiosas para el sistema de justicia penal, su implementación debe realizarse con extrema precaución. Debe desarrollarse un marco regulatorio consistente que equilibre la innovación tecnológica con la protección de los derechos y garantías fundamentales del proceso penal. No debe olvidarse que la gran responsabilidad de los juristas debe basarse en la prudencia a la hora de poner límites a las actuaciones de la ciencia y la técnica, lo que les exige unas adecuadas pautas precautorias a la hora de regular estos aspectos. Y que es misión de los jueces y tribunales de justicia aplicar e interpretar los resultados que se obtienen con estos avances tecnológicos. La IA no tiene la última palabra.

Por tanto, la peligrosidad del sujeto concreto que está siendo enjuiciado, ha de ser apreciada por tanto por dicho juez, cuya valoración no pude sustituirse por la que le proporcione un sistema técnico que ha obtenido conclusiones a partir de experiencias que pueden ser ajenas a la persona juzgada, que precisamente es la titular del derecho a un juicio justo con todas las garantías. El hombre, en este caso

como juez, está destinado al conocimiento de la verdad, a ser capaz de tomar decisiones y, finalmente, a decidir entre el bien y el mal, entre lo justo y lo injusto. Esta capacidad sin duda pertenece exclusivamente al ser humano.

Es por ello que, los resultados que proporcione la IA tienen que ser fiables, sin sesgos y utilizados como apoyo, sin olvidar que están en juego el honor, la privacidad, la libertad y la dignidad de las personas, por lo que debe existir el compromiso de utilizarlos al servicio de cada persona en concreto, de la humanidad en su conjunto y del bien común con el objeto de mejorar el bienestar y la libertad de los seres humanos. La centralidad de la persona es el único planteamiento posible para una IA ética. La meta debe ser aprovechar las ventajas de la IA para complementar pero no para reemplazar el juicio humano en la administración de justicia.

Por todo ello, la prudencia ha de considerarse un principio deontológico más de las profesiones jurídicas, pues la aplicación de la tecnología como instrumento al servicio de aquella debe ponderarse por el operador jurídico, fomentando la confianza del ciudadano, actuando con diligencia, pericia y exigencia de cualificación profesional en los ámbitos técnico, ético y jurídico. Finalmente los Estatutos de las organizaciones colegiales correspondientes deben delimitar más el uso de la IA, con el fin de establecer las posibles responsabilidades.