UTILIZACIÓN DE SISTEMAS DE INTELIGENCIA ARTIFICIAL POR ADMINISTRACIONES PÚBLICAS: UN SISTEMA PROPIO DE GARANTÍAS COMO REQUISITO IMPRESCINDIBLE PARA SU VIABILIDAD

Por Matilde Carlón Ruiz
Catedrática de Derecho Administrativo. UCM

SUMARIO: I.- Algunas reflexiones introductorias: la Inteligencia Artificial como fenómeno ineludible para el Derecho Administrativo. **II.-** Presente y futuro de la Inteligencia Artificial en el sector público desde la perspectiva de su necesario control. **1.-** Qué es la Inteligencia Artificial: mito y realidad. **A.-** Pluralidad de manifestaciones de una tecnología basada en el dato: la inteligibilidad del proceso en el foco. **B.-** La Inteligencia Artificial no es equiparable a la humana. **2.-** Manifestaciones de la Inteligencia Artificial en la actuación pública: presente y futuro. **3.-** Controles y garantías en la incorporación de Inteligencia Artificial *en la toma de decisiones administrativas*: una propuesta de sistematización en el marco del art. 23 de la Ley 15/2022 en conexión con el Reglamento europeo de Inteligencia Artificial. **III.-** Mecanismos para la minimización de riesgos en el uso de Inteligencia Artificial por Administraciones Públicas conforme al principio del ciclo de vida. **1.-** El más elemental mecanismo de control: la racionalización de la decisión misma de la implantación de un sistema de IA. **A.-** El principio del ciclo de vida como paradigma para el análisis –y gestión- de riesgos ante el espejo del Reglamento Europeo de Protección de Datos: evaluaciones de impacto, bancos de pruebas e informes preceptivos como instrumentos preventivos. **B.-** El impacto directo de la propuesta de RIA en los márgenes de utilización de sistemas de IA por Administraciones Públicas: usos prohibidos y de alto riesgo. **2.-** La introducción de un mecanismo continuado de gestión de riesgos como clave de bóveda del modelo regulatorio conforme al principio del ciclo de vida: especial énfasis en los requisitos de trazabilidad y seguridad en el contexto del Esquema Nacional de Seguridad. **3.-** Mecanismos de garantía en la gestión de los datos frente a los riesgos de vulneración de la privacidad y la igualdad. **A.-** La normativa de protección de datos personales como parámetro de garantía para la utilización de sistemas de IA por Administraciones Públicas: auditorías, certificaciones y evaluaciones en escenarios complejos. **B.-** Especial atención a los sesgos y a sus circunstancias en el seno de la Ley 15/2022. **4.-** Las garantías de vigilancia humana y su proyección sobre la inteligibilidad del sistema en aras de su transparencia. Remisión. **IV.-** Garantías de transparencia a favor de los ciudadanos cuando los sistemas de IA son utilizados por Administraciones Públicas: variantes en el cuándo y el cómo. **1.-** Algunas precisiones sobre el carácter gradual de las exigencias de transparencia en función del ámbito de aplicación de las soluciones de IA por Administraciones públicas. **2.-** Alcance de la obligación de publicidad de la utilización de sistemas de IA por parte de las Administraciones Públicas: el art. 41 LRJPAC como referente. **3.-** Parámetros de transparencia en relación con el funcionamiento *ad intra* de los sistemas de IA para la propia Administración: entre lo deseable y lo posible. **4.-** Parámetros de transparencia en relación con el funcionamiento *ad extra* de los sistemas de IA desde el punto de vista del destinatario: entre lo posible y lo deseable. **V.-** Mecanismos de rendición de cuentas por la utilización de sistemas de IA por Administraciones Públicas. **1.-** El pleno alcance del control judicial de las resoluciones adoptadas con IA: al hombre lo que es del hombre y a la máquina lo que es de la máquina con el régimen de silencio administrativo como referente. **2.-** Mecanismos de supervisión y control: en especial, auditorías y evaluaciones de impacto. Luces y sombras en el papel de la Agencia de Supervisión de la Inteligencia Artificial *avant la lettre*. **V.-** Unas breves reflexiones para concluir

I.- Algunas reflexiones introductorias: la Inteligencia Artificial como fenómeno ineludible para el Derecho Administrativo

La Inteligencia Artificial ha irrumpido con fuerza acelerada en nuestro día a día y se ha convertido en tema de moda en todos los foros. Tendrá, sin duda, que ver en ello la mezcla de fascinación y temor que infunde un fenómeno que como su propia denominación pretende evocar, desde sus primeros desarrollos busca emular –y aun superar- la inteligencia humana[1].

Fascinación y temor son, en efecto, los sentimientos encontrados que infunde una tecnología que en sus muy diversas variantes, desde sus primeros estadios, ha pretendido replicar las capacidades intelectuales del ser humano. Los retos filosóficos y éticos que plantea se manifiestan evidentes y por sí mismos fascinantes. Y en sus coordenadas hay que entroncar los retos para el Derecho en general, y el Derecho administrativo en particular, como obra humana que es en su definición y en su aplicación.

El Derecho administrativo, como Derecho que disciplina el ejercicio del poder para el cumplimiento de los intereses generales garantizando a su vez la posición de los ciudadanos respecto de los que se proyecta, se ve profundamente interpelado por un conjunto de tecnologías que pueden facilitar el cumplimiento eficaz y eficiente de algunas funciones públicas, pero que no pueden sacrificar a tal fin las exigencias de control del poder trabajosamente construidas durante decenios. En el Derecho administrativo se reconoce, en efecto, con especial viveza la tensión implícita en la reacción ante la Inteligencia Artificial: Fascinación por las posibilidades que puede llegar a abrir a una acción pública más hábil para lograr el cumplimiento de los intereses generales con un

---

[1] Los orígenes de la IA se remontan a 1943 pero consistieron sólo en el desarrollo del primer modelo de neurona artificial. El primer desarrollo propiamente dicho se produjo en 1957, fecha en la que se presenta el sistema Perceptrón, que ya incluía capacidad de aprendizaje e intentaba emular la percepción visual. Con todo, probablemente ha sido la irrupción de la tecnología GPT alrededor de 2020 la que ha marcado un nuevo hito de entusiasmo en la evolución de la Inteligencia Artificial, que –conviene advertirlo para templar los excesos de optimismo- ha pasado por estadios sucesivos de euforia y decepción en el largo afán de emular las capacidades humanas. Quede aquí testimonio de agradecimiento a la Profa. Dra. Fernández Chamizo, Catedrática de Lenguajes y Sistemas Informáticos de la UCM, por sus valiosísimas explicaciones y pacientes respuestas a mis preguntas.

uso más eficiente de sus recursos, pero, a la par, temor –o al menos vértigo y desconcierto- por los riesgos que generaría un ejercicio deshumanizado del poder público.

En el Derecho Administrativo se plantea, en fin, con la más plena intensidad, la dicotomía que parece enfrentar a la Inteligencia Artificial con la Inteligencia Humana. Una dicotomía que encierra en sí misma luces y sombras, y está llena de malentendidos y contradicciones. La Inteligencia Artificial parecería capaz de suplir, incluso mejorándolas, las capacidades del razonar humano, con la garantía añadida de la objetividad y la pureza en los tiempos del dataísmo[2], lo que nos conduciría a un ejercicio de las funciones públicas más plenamente acorde con el mandato del 103.1 CE, frente a lo que hay quienes llegan a reivindicar los rasgos de la subjetividad más emocional como requisito para el mejor ejercicio de las potestades públicas. Pero ni la Inteligencia Artificial es tan inteligente, ni es tan pura, como obra humana –y obra inacabada- que es. Ni, por su parte, los seres humanos concretos que encarnan la Administración y ejercen las funciones –formalizadas o no- que el principio de legalidad les legitima –y obliga- a desplegar son seres que, en su deseada empatía, carezcan de sesgos tan o más severos que los que se hace necesario evitar o minimizar en la Inteligencia Artificial por cuanto resultan mucho más incisivos por su capacidad de diseminación, multiplicación e, incluso, convicción. Del mismo modo que tampoco carece de sesgos –conviene tenerlo presente- el juez en el que confiamos para controlar la actuación de la Administración, y ello en el bien entendido de que, en su función, también puede hacerse presente la Inteligencia Artificial.

La Inteligencia Artificial se ofrece, pues, como una ocasión, para pensar y repensar elementos centrales de nuestra disciplina, pues por su propio planteamiento se coloca

[2] Evocamos, evidentemente, a Y. H. HARARI (2016: 400-431), para quien, desde una visión negativa, "el dataismo invierte la pirámide tradicional del conocimiento. Hasta ahora, los datos se veían únicamente como el primer eslabón de una larga cadena de actividad intelectual. Se suponía que los humanos destilaban los datos para obtener información, destilaban la información para obtener conocimiento, y este se destilaba en sabiduría. Sin embargo, los dataistas creen que los humanos ya no pueden hacer frente a los inmensos flujos de datos actuales ni, por consiguiente, destilar los datos en información ni mucho menos en conocimiento o sabiduría. Por lo tanto, el trabajo de procesar los datos debe de encomendarse a algoritmos electrónicos, cuya capacidad excede con mucho a la del cerebro humano. En la práctica, esto significa que los dataistas son escépticos en relación con el conocimiento y la sabiduría humanos, y que prefieren poner su confianza en los datos masivos y los algoritmos informáticos".

potencialmente en el núcleo mismo de la acción administrativa –y su necesario control-. En el plano de las garantías nos corresponde concentrarnos en esta ponencia, indagando sobre los mecanismos necesarios para disciplinar, garantizando la posición de los ciudadanos, un fenómeno que interpela de forma ineludible al Derecho Administrativo, obligándole a interiorizarlo de forma equilibrada. Es esta, sin duda, la de las garantías, una cuestión capital, pues la viabilidad misma de la utilización de soluciones de Inteligencia Artificial por las Administraciones públicas depende radicalmente de los mecanismos que incorpore para embridar, que no eliminar –porque hacerlo es imposible-, los riesgos que su uso comporta.

Siendo este el objetivo, el primer paso es comprender el fenómeno en su heterogeneidad y su complejidad y, de forma inmediata, plantearse hasta qué punto se hace presente actualmente y podría llegar a hacerse presente en el futuro en la actuación de las Administraciones Públicas.

II.- Presente y futuro de la Inteligencia Artificial en el sector público desde la perspectiva de su necesario control

1.- Qué es la Inteligencia Artificial: mito y realidad

A.- Pluralidad de manifestaciones de una tecnología basada en el dato: la inteligibilidad del proceso en el foco

El fenómeno de la Inteligencia Artificial impide reducirlo a una única tecnología y a una única aplicación, ya que incluye múltiples desarrollos que despliegan concretas habilidades cognitivas pretendidamente comparables con las del ser humano con el objetivo de favorecer la automatización de procesos basados en el análisis de información[3]. Esta falta de unidad favorece los equívocos respecto de su alcance e impacto, actual y potencial, particularmente cuando se proyecta sobre la acción administrativa, que es la que debe concentrar nuestra atención.

[3] Merece ser consultado al respecto el clarificador estudio de F. GONZÁLEZ CABANES, y N. DIAZ DIAZ (2023: 38-72), y las precisiones de M. MERCHÁN ARRIBAS (2020: 3-9), así como los apuntes, ponderados, atemperando el triunfalismo, de A. ZLOTNIK (2019: 24-27).

A nuestros efectos, lo más oportuno es acudir a la definición de Inteligencia Artificial –o, más propiamente, de Sistema de Inteligencia Artificial- que ofrece la propuesta de Ley europea sobre Inteligencia Artificial de 2021 -en puridad, Reglamento de Inteligencia Artificial (RIA)- cuya aprobación definitiva parece que está a punto de culminar[4]. Y ello no solo porque esta definición es bastante completa y comprensible, habiendo sido elaborada conforme a un cierto consenso previo[5], sino también porque sobre este concepto pivotan las reglas que, bajo la lógica del control del riesgo –esto es clave-, vendrán a disciplinar la utilización de la Inteligencia Artificial en el futuro cercano, sin perjuicio de las modificaciones a las que la redacción final de lugar. También cuando sus proveedores y usuarios –en los términos en que ambos conceptos se acotan por la propia propuesta de ley- son "autoridades públicas", según la expresión que maneja la propia ley[6], sin perjuicio, evidentemente, de su papel de reguladoras en relación con lo que la

[4] Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión [COM (2021) 206 final, de 21 de abril de 2021; {SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}]. Una aproximación a las claves de la propuesta en E. GAMERO CASADO (2021), mientras que N. NIKOLINAKOS (2023) ofrece un recuento exhaustivo de sus antecedentes, referentes, coordenadas y contenido. El 14 de junio de 2023 el Parlamento Europeo aprobó el texto integrando sus enmiendas, con lo que continúa su tramitación en el trílogo, que, sin embargo, se ha visto ralentizada en las últimas semanas por las tensiones en relación con la ordenación de los Sistemas de Propósito General y de los llamados modelos fundacionales. El viernes 8 de diciembre parece, sin embargo, que se ha llegado a un acuerdo político sobre el texto. En la página sobre la Ley elaborada por el *Future of Life Institute* se puede encontrar información extensiva sobre el proceso y sus coordenadas, siendo especialmente ilustrativo el cuadro comparativo, a fecha 20 de junio de 2023, que se puede consultar en https://artificialintelligenceact.eu/es/documentos/ . En el texto hemos preferido, con todo, en el estadio actual del proceso de aprobación, partir de la propuesta de la Comisión.

[5] Lo que no ha impedido que haya sido ya objeto de algunas críticas, por exceso, y por defecto, tal y como pone de manifiesto el elocuente Informe de los SERVICIOS DEL PARLAMENTO EUROPEO (2023), que remarca la multiplicidad de definiciones que han sido propuestas para abarcar lo que en último término engloba un conjunto de aplicaciones informáticas basadas en distintas técnicas que despliegan capacidades comúnmente asociadas con la inteligencia humana. Muchas de las definiciones antecedentes fueron objeto de un análisis comparativo en SAMOILI et al. (2020). Cumple advertir, con todo, que como pone de manifiesto el informe de los Servicios del Parlamento Europeo, la definición del proyecto de reglamento sigue muy directamente la estela de la definición acuñada por el Consejo de la OCDE (2019). En ello insiste también N. NIKOLINAKOS (2023: 351-369), con referencia exhaustiva al contexto de elaboración de la definición. J. I. CRIADO (2021), por su parte, desde un planteamiento más amplio, da noticia de las distintas iniciativas internacionales preocupadas por disciplinar el fenómeno de la IA con las que se alinea la apuesta europea frente a la aproximación estadounidense o china.

[6] En los términos del apartado 2 del mismo artículo 3 de la propuesta de Reglamento, "proveedor" es "toda persona física o jurídica, *autoridad pública*, agencia u organismo de otra índole que desarrolle un sistema de IA o para el que se haya desarrollado un sistema de IA con

misma concibe primariamente como producto en el contexto del mercado único. Esto es importante y sobre ello volveremos.

En los términos del art. 3.1 de la propuesta, «Sistema de inteligencia artificial (sistema de IA)» es "el software que se desarrolla empleando una o varias de las técnicas y estrategias que figuran en el anexo I y que puede, para un conjunto determinado de objetivos definidos por seres humanos, generar información de salida como contenidos, predicciones, recomendaciones o decisiones que influyan en los entornos con los que interactúa", donde el anexo I menciona tanto estrategias de aprendizaje automático, como las basadas en la lógica y el conocimiento y las estrategias estadísticas, de estimación bayesiana, métodos de búsqueda y optimización[7].

La definición pone énfasis en la participación humana en la fijación de los objetivos -que no necesariamente en la supervisión y validación del proceso de obtención de los resultados, y de los resultados mismos- a que dé lugar la aplicación de las técnicas que se contemplan en el anexo, lo que no deja de ser coherente con su contenido, que abarca tanto las técnicas de *Machine Learning* como aquellas en las que el sistema no llega a actuar de forma independiente en la fijación de instrucciones posteriores a la originalmente fijada por el ser humano. De ahí también que la definición contemple una multiplicidad de posibles resultados en lo que identifica como "información de salida", como puedan ser "contenidos, predicciones, recomendaciones o decisiones que influyan en los entornos con los que interactúa"[8], según fórmula que cobra un especial sentido

vistas a introducirlo en el mercado o ponerlo en servicio con su propio nombre o marca comercial, ya sea de manera remunerada o gratuita", mientras que, en los términos del apartado 4, "Usuario" es "toda persona física o jurídica, *autoridad pública*, agencia u organismo de otra índole que utilice un sistema de IA bajo su propia autoridad, salvo cuando su uso se enmarque en una actividad personal de carácter no profesional".

[7] Tal y como explica la propia Comisión en el apartado 5.2.1 de la Exposición de Motivos de la propuesta, esta definición procura ser lo más tecnológicamente neutra posible y resistir al paso del tiempo lo mejor posible, habida cuenta de la rápida evolución tecnológica y del mercado en relación con la IA, siendo así que el anexo I contiene una lista detallada de las estrategias y técnicas para el desarrollo de la IA que la Comisión deberá ir adaptando a medida que evolucione la tecnología.

[8] Es de destacar que la definición propuesta por el GRUPO INDEPENDIENTE DE EXPERTOS DE ALTO NIVEL SOBRE INTELIGENCIA ARTIFICIAL (2019) en sus Guías éticas para una IA fiable afirma que la Inteligencia Artificial se manifiesta "*decidiendo* la acción o acciones óptimas que deben llevar a cabo para lograr el objetivo establecido".

cuando se proyecta en el ámbito del Derecho administrativo, a poco que se piense, simplemente, en el concepto clásico de acto administrativo.

Aunque la definición no lo haga expreso, la Inteligencia Artificial, en sus múltiples manifestaciones, pivota siempre sobre tres parámetros: datos, algoritmos y capacidad de computación para aplicar estos a aquellos. Aquellos son la materia prima de procesos que solo son posibles –máxime en sus más recientes variantes- con la utilización de grandes capacidades de computación sometidas a indicaciones que vienen fijadas por los algoritmos.

Los datos son determinantes, cada vez más, tanto en su calidad como en su cantidad, en cualesquiera aplicaciones de Inteligencia Artificial. Tenerlo presente pone en escena, desde este mismo momento, que el control del diseño y uso de aplicaciones de Inteligencia Artificial encontrará en la normativa de protección de datos un anclaje relevante, que sin duda ha inspirado la aproximación general al fenómeno basado en el control de riesgos. Sobre estos datos operan los algoritmos, con mayor o menor automatismo, lo que nos enfrenta con las variantes de Inteligencia Artificial disponibles, ya que el propio concepto de algoritmo no es en absoluto unívoco.

En puridad, los algoritmos no son otra cosa que un conjunto de pasos claramente establecidos para resolver un determinado problema[9]. En estos términos, el algoritmo no es necesariamente una fórmula matemática, ni tiene por qué integrarse en un software. Una simple receta de cocina es un algoritmo, como lo es el propio procedimiento

[9] Según la primera acepción del Diccionario de la Real Academia Española, un algoritmo es un conjunto ordenado y finito de operaciones que permite hallar la solución de un problema. Con esta definición es coherente la que ofrece la Comisión de Garantía del derecho de acceso a la información pública de Cataluña, en dictámenes que ya son de común referencia, como el dictado en los asuntos acumulados 123 y 124/2016, relativo al sistema de conformación de los tribunales de la PAU, que, por lo que ahora interesa, conviene traer a colación para constatar que, en sus propios términos, "un algoritmo, como "procedimiento de cálculo que consiste en cumplir un conjunto ordenado y finito de instrucciones con unos datos especificados para llegar a la solución del problema planteado" o "conjunto finito de reglas que, aplicadas de manera ordenada, permiten la resolución sistemática de un problema, el cual se utiliza como punto de partida en programación informática" (…), no deja de ser un tipo de información, expresado habitualmente en lenguaje matemático o informático (a pesar de que los algoritmos también se pueden expresar de otras muchas maneras, incluyendo los diagramas de flujo, el pseudocódigo y el propio lenguaje natural)". A esta misma definición se refiere J. PONCE (2018), ofreciendo muy reveladoras explicaciones sobre todos estos elementos conceptuales y de funcionamiento.

administrativo. Las correspondientes instrucciones o pasos diseñados en el lenguaje natural que integran un algoritmo pueden, en su caso, ser *traducidos* al lenguaje de programación para su uso informático: se plasmarán entonces en lo que se conoce como "código fuente", que se puede definir como el conjunto de instrucciones expresadas en lenguaje informático que guían la ejecución de un programa[10]. Todo programa informático tiene un "código fuente", e incluso puede afirmarse que necesariamente incorpora uno o varios algoritmos, pero no necesariamente algoritmos que integren soluciones de IA[11].

Bajo esta sencilla aproximación, podemos entender que los algoritmos han estado siempre en la base de la Inteligencia Artificial, en la medida en que todas las iniciativas englobables en la misma consisten en la fijación de instrucciones pautadas a las máquinas –los ordenadores- para obtener un determinado objetivo. A partir de ahí, las variantes son múltiples y se explican en función de los objetivos pretendidos y las estrategias sucesivamente diseñadas para su logro, siempre en el afán de emulación de las capacidades humanas. En un primer momento, en los años 70 del pasado siglo, a través del desarrollo de los llamados "sistemas expertos" basados en reglas deductivas; a partir de los años 80, adoptando una perspectiva inductiva basada en casos, pretendiendo bajo una y otra aproximación replicar la forma de razonar del experto analizada a través de lo que se pasó a conocer como "ingeniería del conocimiento". La frustración de las expectativas en su aplicación a sectores como el diagnóstico médico abrió el camino, en un estadio sucesivo, a algoritmos crecientemente capaces de aprender por sí mismos, aun con distintos grados de instrucción humana.

---

[10] En los términos de la resolución del Consejo de Transparencia y Buen Gobierno de 18 de febrero de 2019, dictada en el asunto R/0701/2018, relativa al sistema Bosco para la comprobación de los requisitos para obtener el bono social eléctrico, y sobre el que volveremos, "el código fuente es el archivo o conjunto de archivos que tienen un conjunto de instrucciones muy precisas, basadas en un lenguaje de programación, que se utiliza para poder compilar los diferentes programas informáticos que lo utilizan y se puedan ejecutar sin mayores problemas. Los usuarios pueden usar el software sin mayores preocupaciones gracias a una interfaz gráfica sencilla que se basa en el desarrollo del código fuente. El usuario no necesita saber el lenguaje de programación utilizado para desarrollar un determinado software".

[11] Piénsese, simplemente, en un programa informático que permita calcular las nóminas de un colectivo, aplicando una determinada fórmula matemática, aun con variantes.

Así, se desarrollaron primero, hasta el 2000, algoritmos que ahora se califican de "clásicos", utilizados extensivamente en el *Machine Learning*, capaces de establecer correlaciones a partir de la detección de similitudes entre casos. Con posterioridad, los algoritmos aplicados en el *Deep Learning* pueden llegar a desplegar "razonamientos" independientes que devienen incomprensibles para el ser humano. La disponibilidad masiva –y creciente- de datos, la ostensible mejora de la capacidad de computación y la invención de nuevos algoritmos disparó, en efecto, a partir del año 2000, y especialmente a partir de 2012, el desarrollo de redes neuronales profundas, cuyos "razonamientos" no resultan inteligibles y explicables para el ser humano, aunque hay líneas de investigación abiertas para lograr su explicabilidad[12]. Gráficamente denominados como cajas negras, estos sistemas de Inteligencia Artificial presentan un reto añadido en su aplicación en el ámbito del Derecho administrativo, tal y como habrá ocasión de argumentar[13].

Fundamental es, en todo caso, tener en cuenta que los algoritmos pueden o no someterse a entrenamientos supervisados por humanos. En el primer supuesto, los sistemas de Inteligencia Artificial ofrecen soluciones de clasificación de casos. Frente a ellos, los sistemas sin entrenamiento supervisado –conocidos como *clustering*- ofrecen soluciones de agrupamiento de casos sobre la base de patrones o similitudes. En uno y otro supuesto, los algoritmos ofrecen, a partir de un planteamiento inductivo, soluciones basadas en criterios de probabilidad, y en tal caso se pueden considerar predictivos. Por su parte, los algoritmos basados en reglas pueden tener un comportamiento totalmente determinado, no predictivo o pueden ofrecer una combinación de ambas aproximaciones[14], ofreciendo como ventaja frente a un programa informático tradicional su mayor flexibilidad, ya que permiten mantener el programa, modificando puntualmente las reglas.

---

12 Es significativo que estos estudios se basan en aplicar métodos deductivos a los resultados obtenidos a partir de procedimientos inductivos. Se trata, en definitiva, de responder a la pregunta de qué reglas pueden desprenderse para explicar los hallazgos que ofrece, de forma inexplicable en su razonamiento, incluso para sus desarrolladores, el sistema de Inteligencia Artificial.

13 Los problemas de explicabilidad limitan gravemente las aplicaciones de las soluciones de Inteligencia Artificial en campos en los que de la asunción de las conclusiones alcanzadas se derivan responsabilidades. Así, por ejemplo, en el ámbito del diagnóstico médico. También, bajo lógica equivalente, en el del ejercicio de potestades públicas.

14 Sistemas clásicos basados en reglas, como MYCIN, que se desarrolló a principios de los 70 para diagnóstico médico, tienen también cierto carácter predictivo. Este sistema asociaba a cada regla un factor de certeza y, al combinar reglas para obtener una solución, también combinaba sus factores de certeza para obtener el factor de certeza de las conclusiones.

Resulta, así, coherente que la definición de IA propuesta por la ley europea haga expreso que "la información de salida" pueden ser tanto contenidos como predicciones, recomendaciones o, también, decisiones. Ejemplo paradigmático de los sistemas que proporcionan los primeros –es decir, contenidos- son los que, como el revolucionario ChatGPT, se basan en reconocimiento y generación del lenguaje, ofreciendo de forma automática –y sin aplicar verdaderas capacidades de comprensión- textos verosímiles sobre cualquier cuestión apoyándose en documentos previos –como puedan ser de contenido jurídico, como normas, resoluciones o sentencias dentro del panorama que se ha venido en llamar Legaltech[15]-. Las peculiaridades de estas soluciones "de propósito general" y de los llamados modelos fundacionales, no vinculadas con una concreta finalidad, plantean retos específicos.

Y es que, a la vista de todo lo anterior, es de advertir que los algoritmos y los sistemas que los aplican, en su variedad, no ofrecen soluciones uniformes. Dependerá obviamente del objetivo pretendido la viabilidad misma de un sistema de Inteligencia Artificial y, en su caso, la selección del más adecuado tomando en consideración criterios de eficiencia en punto al consumo de recursos de todo orden y en consideración a los riesgos que su manejo pueda implicar. Esta elemental aclaración enlaza con la cuestión crucial de los usos habidos y por haber, actuales y potenciales, de la Inteligencia Artificial en la órbita de la acción pública, particularmente en España. Pero antes debemos apuntar alguna reflexión sobre la comparabilidad real de la Inteligencia Artificial con la Inteligencia Humana que aspira a imitar.

B.- La Inteligencia Artificial no es equiparable a la humana

Los desafíos éticos, filosóficos y jurídicos que se derivan del empleo de la Inteligencia Artificial se reconducen a una reflexión nuclear: ¿es la Inteligencia Artificial asimilable a la humana?.

---

[15] De cuyas facilidades se hace eco el propio Gobierno de España en https://datos.gob.es/es/blog/como-mejorar-la-eficiencia-del-sector-juridico-legaltech-y-el-analisis-de-datos

Esta pregunta tensiona toda aproximación a las cuestiones que nos ocupan. Pero resulta radicalmente irresoluble, por cuanto qué sea la Inteligencia humana, en puridad, es algo sometido a discusión. El cerebro humano es una caja negra a la que los neurocientíficos apenas están empezando a acceder[16]. Partir de esta constatación permite, superada la paradoja, asumir una aproximación más cabal al fenómeno de la Inteligencia Artificial, particularmente cuando se aplica en el ámbito jurídico. Por cuanto reafirma su carácter instrumental, y no sustitutivo de la inteligencia humana entendida en su completud, lo que no impide que se le puedan encomendar tareas en sustitución de un ser humano, si las ha de desarrollar de forma más eficaz y eficiente.

La pregunta no es, sin embargo, caprichosa, sino que deriva inexorablemente del planteamiento deliberadamente emulador de la inteligencia humana con el que se ha construido tradicionalmente la Inteligencia Artificial. El que esta, en sus primeros estadios, se haya basado exclusivamente en la replicación de concretas capacidades humanas ha propiciado, indudablemente, reflexiones dirigidas a comparar una y otra.

"La inteligencia artificial no puede pensar porque no se le pone la piel de gallina". En estos términos se pronuncia Byung-Chul Han, en clara evocación de su referente Heidegger, para argumentar que a la Inteligencia Artificial le falta la dimensión "afectivo-analógica" necesaria para desplegar el pensamiento, en tanto que el proceso de comprensión, dirigido a captar el mundo en conceptos, se ve *afectado* por él[17]. Este planteamiento parece poner en escena limitaciones *emocionales* en la Inteligencia Artificial, pero –cuando se trata de proyectar su impacto sobre el Derecho administrativo, regido por el principio de legalidad- resulta más importante poner énfasis en que la Inteligencia Artificial, en su estadio actual, agota sus capacidades en la revelación de

---

16 Muy significativas al respecto son las apreciaciones de R. YUSTE (2019).

17 B.-Ch. HAN (2021: 53-61). Siguiendo este planteamiento, Han afirma que "la Inteligencia Artificial no piensa porque nunca está *fuera de sí*. El *espíritu* originariamente *está fuera de sí mismo* o *estremecido*. La Inteligencia Artificial puede *calcular* con rapidez, pero le falta el *espíritu*. Para el cálculo, el estremecimiento solo sería una perturbación" (p. 54). "El pensamiento procede de forma muy diferente a la inteligencia artificial. La totalidad constituye el *marco* inicial a partir del cual se conforman los hechos. El cambio de disposición anímica como cambio de marco es como un cambio de paradigma que da lugar a nuevos hechos. La inteligencia artificial, en cambio, procesa hechos *predeterminados que siguen siendo los mismos*. No puede a darse a sí misma nuevos hechos" (p. 57).

correlaciones, sin llegar a ser capaz de establecer relaciones de causalidad, menos aun de verdadera *comprensión*[18].

En este contexto resultan especialmente reveladoras las apreciaciones de E. J. Larson, cuando pone de manifiesto que la Inteligencia Artificial carece de uno de los tres rasgos de la inferencia que caracterizan a la Inteligencia Humana. Ha sido capaz, sí -así lo hemos descrito anteriormente- de desplegar razonamientos deductivos, en un primer estadio, y razonamientos inductivos, en una fase sucesiva de su evolución, pero no es capaz de desplegar razonamientos abductivos, que son aquellos que, ante un problema, permiten al ser humano "seleccionar a modo de hipótesis una idea que parece verosímil en un contexto de infinitas posibilidades" y, de forma añadida, variar el significado de los hechos de resultas de esa selección explicadora, "de tal manera que, explicitada la hipótesis, los hechos llegan a verse como señales o como pistas que apuntan a la propia hipótesis". "La computación deductiva y la inductiva han permitido, en definitiva, el desarrollo de la inteligencia artificial débil, que vemos en las máquinas que continuamente nos rodean realizando tareas concretas. Pero no llegan a ofrecer una inteligencia artificial general abductiva, como es la inteligencia humana, capaz de realizar cualquier tarea y resolver cualquier problema"[19].

Todos estos planteamientos deben reconducirse, a nuestros efectos, al ámbito de la aplicación de las normas, asumiendo que esta operación no exige siempre y en todo caso

---

[18] "El Big data sugiere un conocimiento absoluto. Las cosas revelan las correlaciones secretas. Todo se vuelva calculable, predecible y controlable. Se anuncia toda una nueva era del saber. En realidad, se trata de una forma de saber bastante primitiva. La *data mining* o minería de datos descubre las correlaciones. Según la lógica de Hegel, la correlación representa la forma más baja del saber. La correlación entre A y B dice: A ocurre a menudo junto con B. Con la correlación no se sabe *por qué* sucede esto. *Simplemente sucede*. La correlación indica probabilidad, no necesidad. Se diferencia de la causalidad, que establece una necesidad: *A causa B*" (pp. 57-58). Aun cuando la relación de causalidad no basta para comprender: "El "concepto" –C- es el que permite captar la conexión entre A y B. "La Inteligencia Artificial nunca alcanza el nivel conceptual del saber. No *comprende* los resultados de sus cálculos. El cálculo se diferencia del pensamiento en que no forma conceptos y no avanza de una conclusión a otra".

[19] Tomamos las referencias de E. J. LARSON (2022) de J.A. VALOR YÉBENES (2023), que ofrece una explicación iluminadora de la cuestión en el contexto del fenómeno del Transhumanismo, con cita de otros autores que han tratado las facultades de abducción. En palabras de A. ZLOTNIK (2022: 25), "ningún sistema de IA actual puede considerarse una Inteligencia Artificial General (IAG) y, seguramente, todavía faltan varias décadas para que sea una realidad". Apuntan también reflexiones en este sentido M. MERCHÁN ARRIBAS (2020: 5) y A. HUERGO (2023: 745-746).

desplegar todas las capacidades del intelecto humano, ni siquiera en los supuestos en los que la operación se aleja de automatismos y se adentra en el plano de lo valorativo. En estos términos, la introducción de la Inteligencia Artificial en el escenario de la aplicación del Derecho ofrece una nueva oportunidad para volver sobre la clásica reflexión acerca de los parámetros dentro de los que aquella debe producirse: como resultado de un razonamiento humano pleno o, en el extremo opuesto, como fruto de un automatismo aséptico en el acto de aplicar la norma. Recordemos la histórica aspiración de Montesquieu del juez reducido a ser "una boca que pronuncia las palabras de la ley", en expresión máxima de los principios de legalidad y de división de poderes enraizada en una visión racionalista hasta el extremo.

No cabe duda de que semejante aproximación ha quedado plenamente superada, lo que no impide reconocer que en la aplicación normativa, en función de la estructura y contenido de las normas a aplicar, el grado de libertad de su aplicador puede variar considerablemente, sin que quepa descartar supuestos –aun los menos- de virtual automatismo en el proceso. Planteamiento este que encuentra especial correlación, en el ámbito del Derecho Administrativo, en la distinción entre potestades discrecionales y regladas. Ello supone que el margen que en estas últimas pueda jugar la aplicación de sistemas informáticos y, en su caso, de Inteligencia Artificial para la toma de decisiones, será, en principio, potencialmente mucho mayor que el que tolere el ejercicio de potestades discrecionales.

La posibilidad de adoptar, mediante sistemas de Inteligencia Artificial, las correspondientes decisiones viene condicionada, a día de hoy, por las limitaciones mismas de la tecnología, desde un doble plano: el del limitado alcance de su capacidad actual para desplegar según qué razonamientos, y el de su fiabilidad, cuestión esta última que redunda en la necesidad de establecer un régimen de garantías que respalde su viabilidad. Pero en la medida en que –particularmente- aquellas puedan llegar a quedar superadas, la cuestión nuclear que se suscita es si el ser humano, encarnando el correspondiente órgano administrativo, puede ser desplazado por una máquina en el ejercicio de una potestad que requiere una toma de decisión *discrecional*. Ahí estaría el verdadero salto cualitativo.

Cuestiones como la conveniencia –o no- de excluir la aplicación de sistemas de IA en determinadas actuaciones administrativas, la exigencia –o no- de supervisión humana de los sistemas –durante el proceso y/o para validar el resultado- y de su explicabilidad para la propia Administración que los aplica y para los ciudadanos que son destinatarios de sus resultados, se revelan, pues, desde este momento, claves. Sobre ellas reflexionaremos oportunamente en los sucesivos apartados de esta ponencia, pero para hacerlo de forma cabal debemos abordar primero la aproximación prometida sobre el alcance actual y potencial de aplicación de sistemas de IA en la órbita de acción de las Administraciones españolas, al hilo de la cual avanzaremos alguna opinión al respecto.

Interesa, con todo, tener presente desde ahora una última apreciación como colofón de este epígrafe. Son las carencias de racionalidad, y no las de emocionalidad, en los sistemas de Inteligencia Artificial, las que deben justificar, en su caso, la reserva de humanidad que se ha reivindicado en el manejo de tales sistemas en el ámbito de la actuación pública. Y ello como consecuencia última del principio de legalidad que ahorma todo el quehacer de la Administración, que impone que los márgenes de la empatía –que se ha llegado a reivindicar como argumento de aquella reserva- no puedan ser otros que los que la propia ley haya dispuesto[20].

2.- Manifestaciones de la Inteligencia Artificial en la actuación pública: presente y futuro

La Inteligencia Artificial, en la gran amplitud y diversidad de aplicaciones y funcionalidades que abarca y aun dentro de las limitaciones que su actual estadio de evolución impone, puede ofrecer a las Administraciones Públicas oportunidades para un cumplimiento más objetivo y, en último término, más eficaz y eficiente de los intereses generales. En estos términos, puede resultar un instrumento valioso para cumplir los objetivos que les señala el art. 103.1 CE, siempre y cuando ello no suponga un riesgo para el pleno cumplimiento del principio de legalidad y para los derechos de los particulares,

[20] Como es bien sabido, particularmente J. PONCE (2019) y (2022), defiende ese requerimiento de empatía en el ejercicio de potestades discrecionales. Sobre ello habrá ocasión de volver.

especialmente en la esfera de sus derechos fundamentales[21]. Bajo tan elementales parámetros hay que abordar el posible uso de técnicas de Inteligencia Artificial por las Administraciones Públicas.

Conviene advertir, sin embargo, desde este mismo momento, que la diversidad de manifestaciones que presenta la Inteligencia Artificial favorece que su utilización por las Administraciones Públicas encuentre escenario, tanto en relación con actuaciones no necesariamente procedimentalizadas, y en todo caso no integradas en un procedimiento administrativo en sentido propio, como en verdaderos procedimientos administrativos dirigidos a la adopción de actos en ejercicio de potestades administrativas, particularmente de contenido decisorio, siendo así que en este segundo caso puede limitarse a ofrecer elementos para la toma de decisión o puede llegar incluso a determinar la decisión a tomar[22]. El impacto del uso de Inteligencia Artificial en esta segunda esfera se demuestra, evidentemente, con mayor intensidad y presenta rasgos genuinos respecto de los que son comunes al uso de esta tecnología por particulares. De ello se derivarán, consecuentemente, exigencias más intensas cuando de lo que se trata es de determinar las garantías que necesariamente han de articularse para su utilización, hasta el punto de que estas garantías pasarán de ser un elemento accesorio para constituirse en un elemento consustancial a la legitimidad de la utilización de sistemas de IA en la toma de decisiones.

Es esta, sin duda, una cuestión crucial. La implantación de las debidas garantías en el funcionamiento de las soluciones de IA es la que hará verosímil su aplicación, incluso, en la esfera de la toma de decisiones, sin descartar necesariamente las de contenido discrecional. Y ello en el bien entendido de que –sin incurrir en ingenuidades- el alcance

---

[21] C. TORRECILLA-SALINAS et al. (2023: 76-78) glosa las ventajas que puede suponer la introducción de funcionalidades de Inteligencia Artificial en el sector público, sin ocultar sus posibles riesgos, generales y específicos.

[22] M. MERCHÁN ARRIBAS (2020: 13) enuncia las siguientes aplicaciones de soluciones de IA en el sector público: "La automatización de tareas repetitivas de los procedimientos que consumen tiempo, liberando a los empleados públicos de esas tareas (tecnología softbots RPA); la mejora de las decisiones públicas proporcionando información más precisa, pronósticos y predicciones que conduzcan a mejores resultados (aprendizaje automático); la mejora de los servicios públicos mediante el uso de IA para proporcionar soluciones más cercanas y personalizadas (chatbots); la simulación de sistemas complejos que permiten experimentar con diferentes opciones y detectar consecuencias no deseadas antes de tomar decisiones; el reconocimiento de imágenes para agricultura, seguridad pública, etc. (Redes Neuronales Artificiales)". Todo el documento es muy ilustrativo de distintas iniciativas.

real de tales garantías difícilmente puede ser absoluto en algunos sistemas, dada su propia configuración, muy en particular cuando se trate de redes neuronales.

Conviene tener todo esto bien presente desde este momento, aun cuando sea necesario advertir que en el estadio actual de desarrollo de la Inteligencia Artificial, los supuestos de uso de soluciones de IA en el sector público se concentran, tanto en el ámbito nacional como en el europeo, en la primera esfera de las actuaciones no integradas en procedimientos administrativos.

Los dos estudios llevados a cabo en los últimos años por el JRC de la Comisión Europea dentro de su proyecto "AI Watch" son bien ilustrativos de un panorama modesto, pero en crecimiento, en la utilización de tecnologías de Inteligencia Artificial por las Administraciones nacionales[23], en el que son significativamente escasos los supuestos de utilización en procedimientos propiamente decisorios[24]. En ambos estudios se da cuenta de algunas iniciativas españolas, muy heterogéneas, mayoritariamente relacionadas con actividades auxiliares, aunque no necesariamente exentas de impacto, como es el caso de supuestos bien conocidos como el sistema VioGen, para calibrar el riesgo de que mujeres puedan sufrir violencia machista; el Veripol, para la detección de falsas denuncias policiales; el sistema de estimación de ingresos de los trabajadores autónomos para el pago de sus tributos por módulos[25]; otros que facilitan el triaje de enfermos, como en el del servicio de psicología del Hospital San Carlos; o el sistema Corpus Viewer que utilizan los Ministerios de Economía y Ciencia para definir las áreas de actuación en las

[23] El estudio de J. I. CRIADO (2021) pone de manifiesto lo incipiente de la aplicación de la IA por las Administraciones Públicas, resultando prematuro evaluar su eficacia con vistas a orientar sus aplicaciones futuras. El estudio comparado de WOLSWINKEL, J. (2022), siendo modesto en su objeto, ofrece algunas referencias de interés.

[24] Si en el primer informe se analizan 8 casos, en el segundo se da cuenta de 686, de los que solo el 38% están plenamente implantados, pudiendo ser consultados en https://ipsoeu.github.io/ips-explorer/. C. TORRECILLA-SALINAS et al. (2023: 80-87) dan cuenta de ambos informes, extensivamente del segundo, concluyendo que "actualmente el uso de la IA en los servicios públicos se centra mucho más en la automatización de tareas y la realización de predicciones que en la toma de decisiones automáticas".

[25] Este caso es estudiado, junto con solo otros 7 de otros Estados, por L. TANGI et al (2022). Se destaca, con todo, que sus resultados se manejan a título orientativo, completándose con otras comprobaciones. También G. MISURACA y C. VAN NOORDT (2020: 47) se hacen eco de este caso, por más que no ofrezcan una visión general de la cuestión en España.

políticas de investigación[26]. Estos estudios no dan cuenta, sin embargo, de otras iniciativas como el sistema de alerta anticorrupción SALER implantado en la Comunidad Valenciana en 2018, aun con severas advertencias de la AEPD, y del que se ha hecho eco la doctrina[27].

En otras ocasiones, extensivamente en el ámbito local, los sistemas de IA son empleados para interactuar con los ciudadanos en servicios de atención e información, algo de lo que podemos dar fe, en primera persona, cualquiera de nosotros[28].

Son muy escasos los supuestos en los que soluciones de IA se apliquen adentrándose en el espacio de la toma de decisiones[29]. Aunque, como bien precisa la Agencia Española de Protección de Datos, "no todo sistema que toma una decisión automatizada es IA, no toda IA es *Machine Learning*, ni todo lo que se publicita como IA es realmente IA"[30]. Resulta, de hecho, dudoso que sea un supuesto de IA el muchas veces citado sistema Bosco que aplica el Ministerio de Energía para el reconocimiento de la condición de usuario vulnerable a los efectos de obtener el bono social, que ha dado lugar a un pleito de interés, sobre el que volveremos[31]. Como lo es también el caso del algoritmo Euphemia, que casa

---

[26] En la página del Ministerio de Economía se da cuenta de los pormenores del sistema, según consta en https://plantl.mineco.gob.es/tecnologias-lenguaje/actividades/plataformas/Paginas/corpus-viewer.aspx

[27] El sistema, introducido mediante Ley 22/2018, de 6 de noviembre, de Inspección General de Servicios y del sistema de alertas para la prevención de malas prácticas en la Administración de la Generalitat y su sector público instrumental, fue objeto, en su Anteproyecto de Ley, de un severo informe de la AEPD del que da cuenta L. COTINO HUESO (2021). Una descripción del mismo aparece en la página del Centro para al Conocimiento Antifraude de la Comisión Europea: https://antifraud-knowledge-centre.ec.europa.eu/library-good-practices-and-case-studies/good-practices/saler-rapid-alert-system_es

[28] Un paso más allá vendría dado por la implantación de sistemas de prestación electrónica de servicios proactivos, siguiendo experiencias como la estonia, para lo que se requiere un grado de madurez de la Administración electrónica aun no alcanzado en nuestro país. Al respecto, C. I. VELASCO RICO (2020: *in totum*).

[29] Es especialmente llamativo un caso –denominado Robocop- que el Informe atribuye a España, aunque leído atentamente, parece referirse a Hungría, y conforme al cual se aplicaría automáticamente un sistema sancionatorio de tráfico, a partir de las denuncias constatadas por las cámaras de detección de infracciones.

[30] AEPD (2020: 7).

[31] Resuelto por SAN de 30 de diciembre de 2021 (Roj: SAN 5863/2021), en la que se respaldó, como veremos, la denegación de acceso al código fuente del sistema decidida por el Consejo de Transparencia y Buen Gobierno en su resolución de 18 de febrero de 2021 (R/0701/2018).

los precios del mercado eléctrico[32]. Otra cosa es el alcance que pueda tener el uso de Inteligencia Artificial por parte de la propia CNMC a través de su Unidad de Inteligencia Económica, cuya finalidad confesa es la detección de oficio de prácticas anticompetitivas o el apoyo para la toma de decisiones de la Dirección de Competencia, incluyendo la detección y análisis de prácticas de colusión algorítmica[33].

La apuesta decidida desde la Estrategia Nacional de Inteligencia Artificial (ENIA) auspiciada por el Gobierno para extender el uso de estas tecnologías en el sector público –aunque por momentos se formule de forma un tanto evanescente[34]- permite pronosticar una extensión del uso de la IA por las Administraciones Públicas para la que, a día de hoy, no hay exclusiones ni límites específicos en nuestro ordenamiento –a salvo algunas previsiones autonómicas relativas a las actuaciones automatizadas-, como apenas hay ordenación ninguna. En este escenario, lleno de dificultades, es en el que hay que incardinar una propuesta de sistematización de los mecanismos, imprescindibles, de garantía de los intereses públicos y los derechos de los ciudadanos frente al reto que

---

[32] El algoritmo de casación Euphemia, según determina la Resolución de la CNMC de 6 de mayo de 2021, por la que se aprueban las reglas de funcionamiento de los mercados diario e intradiario de energía eléctrica para su adaptación de los límites de oferta a los límites de casación europeos (BOE nº 120 de 20 de mayo de 2021), "busca la optimización del denominado «welfare», que corresponde a la suma para el conjunto de todos los periodos horarios del horizonte de programación del beneficio de las ofertas de compra, más el beneficio de las ofertas de venta, más la renta de congestión. Se entiende por beneficio de las ofertas de compra la diferencia entre el precio de la oferta de compra casada y el precio marginal recibido, y se entiende por beneficio de las ofertas de venta la diferencia entre el precio marginal recibido y el precio de oferta de venta casado" (Regla 42.2).

[33] Una aproximación a sus labores se encuentra en https://www.cnmc.es/ambitos-de-actuacion/competencia/unidad-de-inteligencia-economica Sobre el incipiente fenómeno de la colusión algorítmica, vid. C. ROBLES MARTÍN-LABORDA (2018).

[34] La Estrategia fue aprobada en diciembre de 2020 y se refiere a la extensión de la IA en el sector público en el Eje 5, que se puede consultar en sus páginas 56 a 63. En su página 59 afirma como ventajas de la introducción de IA en la actuación de las Administraciones Públicas: "Aumentar la eficiencia de los procesos, con el fin de agilizar trámites, automatizar procesos mediante la implantación de robots, permitir una mejor interrelación con la ciudadanía a través de asistentes virtuales o chatbots, reforzar la seguridad, luchar contra el fraude con modelos basados en detección de patrones fraudulentos, y en general, mejorar la calidad de las políticas públicas con base analítica que permitan obtener políticas óptimas basadas en simulación. Los distintos Ministerios están poniendo en marcha programas para la integración de la IA en sus sistemas con el fin de mejorar sus repositorios de datos y mejorar el servicio a la ciudadanía". Referente de la ENIA es, sin duda, también a este respecto, el Libro Blanco de la UE sobre la inteligencia artificial: un enfoque europeo orientado a la excelencia y a la confianza, Comunicación de la Comisión Europea [COM(2020) 65 final, de 19 de febrero].

implica la implantación de soluciones de IA para la actuación pública en sus muy diversas variantes, más potenciales que reales –a día de hoy- como acabamos de ver.

3.- Controles y garantías en la incorporación de Inteligencia Artificial *en la toma de decisiones administrativas*: una propuesta de sistematización en el marco del art. 23 de la Ley 15/2022 en conexión con el Reglamento europeo de Inteligencia Artificial

A la vista de todo lo anterior, se hace evidente que todo intento de avanzar, en el momento presente, en un intento de explicación sistematizada de las garantías y controles necesarios para disciplinar el uso de la Inteligencia Artificial en el ámbito de la actuación pública se topa con múltiples dificultades. La inmadurez y falta de univocidad del fenómeno en el momento actual impiden, por sí mismas, dar una respuesta homogénea a la cuestión, a lo que se suma la multiplicidad de riesgos que su utilización, en mayor o menos medida, revela. Como obstáculo añadido hay que contar con la endeblez de las fuentes disponibles para el análisis: la práctica inexistencia de norma alguna en el ordenamiento español y la condición de proyecto, en el momento actual, de la que se ha venido en llamar Ley de Inteligencia artificial gestada por la Comisión Europea, cuya aprobación se ha ralentizado.

Bajo estas coordenadas, parece que el más verosímil anclaje para abordar cualquier propuesta de sistematización de los controles previstos y deseables a proyectar sobre las distintas manifestaciones de la Inteligencia Artificial en la órbita de la actuación de las Administraciones públicas en el ordenamiento español se encuentra en la que hoy por hoy es la única norma que en la legislación estatal aborda el fenómeno: el art. 23 de la Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación. Este artículo, integrado en una ley cuyo objeto trasciende al fenómeno que nos ocupa –lo que, en todo caso, condiciona su contenido-, ofrece –aun con sus deficiencias- unos parámetros útiles para articular el esfuerzo que nos corresponde hacer. El artículo acierta, además, al formular el fenómeno de la Inteligencia Artificial en el ámbito del sector público por referencia a "*los algoritmos involucrados en la toma de decisiones por parte de las Administraciones Públicas*": no tanto porque, como sabemos y ampliaremos, sea exclusivamente sobre los algoritmos empleados sobre los que deban girar las garantías de cualesquiera sistemas de IA, como porque engloba ampliamente tanto supuestos de

utilización de IA *para* la toma de decisiones como *en* los elementos preparatorios o accesorios del proceso decisorio[35].

El apartado 1 del artículo que nos ocupa, bajo el enunciado "Inteligencia Artificial y mecanismos de toma de decisión automatizados", se pronuncia en los siguientes términos:

> "En el marco de la Estrategia Nacional de Inteligencia Artificial, de la Carta de Derechos Digitales[36] y de las iniciativas europeas en torno a la Inteligencia Artificial, las administraciones públicas favorecerán la puesta en marcha de mecanismos para que los algoritmos involucrados en la toma de decisiones que se utilicen en las administraciones públicas tengan en cuenta criterios de minimización de sesgos, transparencia y rendición de cuentas, siempre que sea factible técnicamente. En estos mecanismos se incluirán su diseño y datos de entrenamiento, y abordarán su potencial impacto discriminatorio. Para lograr este fin, se promoverá la realización de evaluaciones de impacto que determinen el posible sesgo discriminatorio".

El precepto se pronuncia en términos no imperativos, pero elocuentes. Compromete, en primer lugar, a las Administraciones Públicas a que favorezcan la puesta en marcha de "mecanismos" para que los algoritmos que utilicen en su ámbito para la toma de decisiones tengan en cuenta un triple parámetro: criterios de minimización de sesgos, transparencia y rendición de cuentas, aun de forma condicionada –introduciendo una buena dosis de realismo- a que todo ello "sea factible técnicamente". Aun con esta precisión de radical importancia, identifica así el precepto tres áreas de atención que resultan perfectamente útiles para abordar una sistematización de las variadas cuestiones que resultan relevantes para asegurar una "Inteligencia Artificial fiable", en expresión deliberada para conectar con la aproximación europea a la cuestión sobre la que volveremos inmediatamente.

Criterios de minimización de sesgos, Transparencia y Rendición de cuentas sirven, en efecto, como enunciados para presentar de forma sistematizada distintas cuestiones respecto de las que el ordenamiento jurídico debe ofrecer garantías –no exclusivamente volcadas sobre los algoritmos-, y muy particularmente el ordenamiento jurídico-administrativo cuando son las Administraciones Públicas las que hacen uso de soluciones

---

35 Ambas posibilidades están presentes en la dicción del artículo, por más que en su título se haga referencia exclusivamente a los "mecanismos de toma de decisión automatizados".

36 La Carta de Derechos Digitales, como es sabido, es un documento de carácter no normativo, elaborado por un grupo de expertos en el marco del Plan de Recuperación, Transformación y Resiliencia.

de Inteligencia Artificial. Criterios de minimización de riesgos, diríamos más ampliamente para formular el primer enunciado, pues los sesgos son unos entre varios riesgos que puede generar la utilización de Inteligencia Artificial, especialmente por Administraciones Públicas.

Esta ampliación del primer enunciado se compadece, además, de forma más fiel con la necesaria puesta en escena de la ya mencionada Ley europea de Inteligencia Artificial a la que apela el artículo que nos ocupa al engarzarse deliberadamente con las iniciativas europeas en la materia. Y ello por cuanto esta norma, que vendrá a disciplinar el uso de la IA tanto por el sector público como por el privado, parte de un análisis de los riesgos que implican las distintas tecnologías englobadas en la Inteligencia Artificial en función del ámbito en que se apliquen bajo un planteamiento que reconoce aquella como un producto protegido por la libertad de circulación en el mercado interior[37].

Bajo esta lógica, y como consecuencia de análisis de riesgos llevado a cabo para la elaboración de la norma, y en función de un criterio de proporcionalidad[38], el proyecto de ley europea prohíbe directamente determinadas aplicaciones de Inteligencia Artificial y somete a supervisión otras que identifica como de Alto Riesgo, a las que se imponen severas obligaciones que se concentran en los elementos más críticos derivados de la utilización de la IA siguiendo planteamientos bien asentados, particularmente en las "Guías Éticas para una IA fiable" elaboradas en 2019 por el Grupo Independiente de

---

37 De ahí que, para cumplir conjuntamente con el fin de lograr un "ecosistema de confianza" en la implantación de la IA en Europa, la propuesta de RIA venga acompañada de una Directiva sobre responsabilidad en materia de IA [Propuesta de Directiva del Parlamento Europeo y del Consejo relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial, COM (2022) 496 final, de 28.9.2022] y una propuesta de revisión de la Directiva sobre responsabilidad por los daños causados por productos defectuosos que incorpora las peculiaridades derivadas del uso de IA [Propuesta de Directiva del Parlamento Europeo y del Consejo sobre responsabilidad por los daños causados por productos defectuosos, [COM (2022) 495 final, de 28.9.2022]. De todos estos instrumentos ofrecen una síntesis elocuente E. GÓMEZ y al. (2023: 732-740).

38 Así se expresa elocuentemente el cdo. 14 de la norma: "Con el fin de introducir un conjunto proporcionado y eficaz de normas vinculantes para los sistemas de IA, es preciso aplicar un enfoque basado en los riesgos claramente definido, que adapte el tipo de las normas y su contenido a la intensidad y el alcance de los riesgos que puedan generar los sistemas de IA en cuestión. Por consiguiente, es necesario prohibir determinadas prácticas de inteligencia artificial, definir los requisitos que deben cumplir los sistemas de IA de alto riesgo y las obligaciones aplicables a los operadores pertinentes, e imponer obligaciones de transparencia a determinados sistemas de IA".

Expertos de Alto Nivel sobre Inteligencia Artificial[39]. Para las demás aplicaciones se incentiva el alineamiento con las exigencias de la norma a través de la suscripción voluntaria de códigos de conducta por parte de los proveedores de IA.

Nos encontramos, así, con un ejemplo paradigmático de regulación basada en los riesgos, que intenta disciplinar un fenómeno que, desde una perspectiva pura de mercado –en su dimensión de producto-, debe estar revestido de garantías para evitar los recelos que de otro modo pudieran lastrar su desarrollo[40]. Unos recelos que entroncan con los derechos de los ciudadanos, en cuya defensa las instancias europeas y las nacionales que han venido enfrentando los retos que plantea el fenómeno se empeñan con convicción íntima, sin que ello resulte contradictorio –si no, al contrario, se retroalimenta- con las razones para asegurar que el mercado interior no se vea lastrado por la aparición de medidas nacionales fragmentarias. Las garantías y controles a establecer buscan proteger a los ciudadanos en cuanto que sujetos involucrados en los sistemas de IA: en su propio diseño y producción, particularmente a través del manejo de sus datos personales, y en cuanto puedan ser usuarios y destinatarios de usos de los mismos, sea como consumidores o en sus relaciones con las Administraciones Públicas. Y al mismo tiempo, buscan proteger el mercado interior y favorecer la industria europea de la IA, siendo para ello crucial introducir garantías y controles que revistan al producto de fiabilidad.

Otra cosa es la relatividad de las posibilidades de un control efectivo, según apuntábamos y desarrollaremos, dadas las limitaciones técnicas intrínsecas a muchos de los sistemas de IA. Cobra, así, pleno sentido la precisión "siempre que sea técnicamente factible" con la que matiza el legislador su llamada a introducir mecanismos de garantía. Y así quedan explicadas y justificadas las expresiones de relativa imprecisión con la que el propio RIA desgrana las múltiples obligaciones que impone a los proveedores de sistemas de IA.

---

[39] En ello pone énfasis N. NIKOLINAKOS (2023: 23-306). En estas Guías, a las hicimos referencia en la nota 8, se identifican 7 Requisitos Éticos Clave, que resultan, como veremos, en mayor o menor medida, reconocibles en el texto de la propuesta de RIA: Acción y supervisión humanas; Robustez técnica y seguridad; Privacidad y protección de datos; Transparencia; Diversidad, no-discriminación y justicia; Bienestar social y medioambiental y Responsabilidad. Ya los había hecho suyos, de hecho, la Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social y al Comité de las Regiones "Generar confianza en la Inteligencia Artificial centrada en el ser humano" [COM (2019)168, de 8 de abril de 2019] y en ellos redunda el posterior Libro Blanco de la UE, COMISIÓN EUROPEA (2020), citado supra.

[40] J. ESTEVE PARDO (2023:167-168), de hecho, incluye una somera referencia a la regulación europea de la IA como ejemplo de Derecho regulatorio de riesgos.

Todo ello no debe suponer una renuncia a las garantías, sino una llamada al realismo frente a la aspiración de domeñar unos riesgos que, cuando la técnica está en juego –sobremanera en el mundo de la informática-, son intrínsecos a una realidad que hace tiempo que ha cobrado vida propia. Es más. Estas garantías, por limitadas que sean, cuando se proyectan sobre el *proceso* de razonamiento de la máquina, ofrecen márgenes de control inalcanzables respecto del razonamiento humano, tal y como tendremos ocasión ponderar.

Resulta, por todo ello, imprescindible leer el art. 23 de la Ley 15/2022 a la luz de las previsiones del proyecto de ley europea para abordar una sistematización de los distintos "mecanismos" –utilizando la expresión de aquel- que deben introducirse como garantías para asegurar una utilización de aplicaciones de Inteligencia Artificial por parte de las Administraciones Públicas que sea respetuosa con las exigencias, cualificadas, del principio de legalidad, el respeto a los derechos fundamentales y de la posición jurídica de los ciudadanos. En unos casos por el mero hecho de que, tratándose de aplicaciones consideradas de Alto Riesgo por el Reglamento europeo, sus exigencias se imponen. En otros, por vías –particularmente normativas o contractuales- mediante las que las Administraciones Públicas puedan imponer o, en su caso, incentivar la utilización de tecnologías de IA ambiciosas en el cumplimiento de exigencias rigurosas, siempre de forma acompasada con el tipo de aplicación en presencia y sin desconocer criterios de proporcionalidad.

Así pues, en los sucesivos apartados abordaremos los distintos mecanismos de garantía y control del uso de la IA por Administraciones Públicas sistematizándolos a partir de la triple categoría apuntada: Minimización de riesgos, Transparencia y Rendición de Cuentas. Y ello en el bien entendido de que estos mecanismos se presentan de forma entrelazada, de modo que esta estructuración sistemática que se propone, aunque tome deliberadamente como base el contenido del art. 23.1 de la Ley 15/2022, tiene no poco de convencional. En ello pesa sobremanera la propia lógica de la regulación de riesgos, en la que las fases de valoración y de gestión de riesgos se plantean de forma circular conforme al principio del ciclo, integrando la de la responsabilidad o rendición de cuentas en este mismo flujo de eterno retorno.

III.- Mecanismos para la minimización de riesgos en el uso de Inteligencia Artificial por Administraciones Públicas conforme al principio de ciclo de vida

Ya hemos puesto de manifiesto que los sistemas de Inteligencia Artificial, en su diversidad, implican riesgos de mayor o menor intensidad, por sí mismos o en función del contexto en el que se utilicen. Tener esto presente en el ámbito de las actuaciones públicas es crucial. Ello supone que el objetivo de minimización de los potenciales riesgos debe analizarse caso a caso, con carácter previo y sucesivo a la propia implantación de un sistema de Inteligencia Artificial, teniendo en cuenta, al respecto, amén de otras consideraciones, los mecanismos que pone en escena el propio RIA y, en conexión con él, la legislación de protección de datos, para minimizar los riesgos derivados de la utilización de estos –a efectos de garantizar su calidad y reducir los potenciales sesgos- y para asegurar la vigilancia humana en el diseño y aplicación de los sistemas.

Hay que tener en cuenta que toda aproximación consustancial a la regulación de riesgos pasa por un análisis y gestión de los riesgos conforme al principio del ciclo de vida, asumiendo que es un proceso continuo que no se agota, aunque encuentra un escenario crítico, en la decisión misma de puesta en servicio del sistema. Esta es la aproximación que hacen, tanto el RIA como la legislación de protección de datos, que ofrece instrumentos muy valiosos, como las auditorías y las evaluaciones de impacto, que han de ser útiles para disciplinar los sistemas de IA incluso cuando no estén directamente en juego datos personales.

Todas estas cuestiones deben ser objeto de análisis diferenciado, con independencia de su íntima imbricación.

1.- El más elemental mecanismo de control: la racionalización de la decisión misma de la implantación de un sistema de IA

A.- El principio del ciclo de vida como paradigma para el análisis –y gestión- de riesgos ante el espejo del Reglamento Europeo de Protección de Datos: evaluaciones de impacto, bancos de pruebas e informes preceptivos como instrumentos preventivos

Entre los mecanismos de minimización de los riesgos que pueda implicar la implantación de un sistema de Inteligencia Artificial en la órbita de la actuación de las Administraciones públicas, el primero y más elemental es el que se concreta en la racionalización de la decisión sobre el si y el cómo de utilizar una concreta aplicación de IA para una concreta acción pública.

Tal y como poníamos de manifiesto al comenzar este estudio, la Inteligencia Artificial está de moda, lo cual supone por sí mismo un riesgo si se tiende a implantar sistemas de Inteligencia Artificial que no resulten, desde un punto de vista técnico y organizativo, ni verosímiles ni eficaces para el cumplimiento de los concretos objetivos pretendidos, ni sostenibles ni eficientes, en términos estrictamente socioeconómicos –incluyendo criterios medioambientales[41]- o, desde una dimensión más amplia, introduciendo consideraciones relativas a los concretos riesgos que pudieran implicar para los derechos de los ciudadanos.

Resultan especialmente pertinentes a este respecto las apreciaciones de la AEPD sobre la base de la constatación de los no pocos riesgos que pueden entrañar desde la perspectiva de la protección de datos[42]: "La disponibilidad o novedad de una tecnología no justifica, por sí misma, su utilización, sino que debe ser objeto de ponderación, realizando un análisis de si el tratamiento, en la forma en la que se plantea, es equilibrado porque se derivan más beneficios y ventajas concretas para el interés general y la sociedad en su conjunto que perjuicios, entendidos estos como los riesgos sobre los derechos y libertades de los sujetos cuyos datos son objeto de tratamiento".

De lo que se trata, en definitiva, es de asegurarse de que cualesquiera proyectos de Inteligencia Artificial que aborden las Administraciones públicas resultan verdaderamente necesarios y oportunos, en la medida en que la incorporación de una solución de Inteligencia Artificial tenga sentido para mejorar el logro de los objetivos

---

41 Es ya un lugar común poner de manifiesto los enormes recursos de energía e, incluso, de agua (para la refrigeración) que requieren los masivos sistemas de computación que dan soporte a algunas de las soluciones de IA. Sirva frente a ello, sin negarnos una nota de humor, la precisión totalmente veraz que hace usualmente R. YUSTE al recordar que el cerebro humano, teniendo como tiene tres veces más neuronas y conexiones que nodos en todo Internet, puede ser alimentado con un simple bocadillo, equivalente a una bombilla de 20 vatios.

42 AEPD (2020: 44).

propuestos, integrando íntimamente un análisis veraz de los riesgos que pueda representar conforme a la consideración de su ciclo de vida completo. Este criterio es clave.

La idea del ciclo de vida del sistema –un concepto común a los desarrollos tecnológicos- es, en efecto, clave, por cuanto se trata, por concepto, de intentar disciplinar unos riesgos que se han de pronosticar y calibrar para el diseño mismo del sistema con el objetivo de decidir su implantación o no, y que, en cuanto tal, está a expensas de los resultados que arroje su aplicación práctica, que puede confirmar o descartar, en más o en menos, los cálculos originales, con las consecuencias que en cada caso correspondan, lo que obliga a implantar medidas oportunas para la gestión de los riesgos previstos e imprevistos. Bajo este planteamiento -que es consustancial, como advertíamos, a la lógica de la regulación de riesgos[43]-, la Administración queda interpelada en tres planos cuando se trata de disciplinar el uso de soluciones de IA para la actuación administradora: como reguladora, en la medida en que pueda decidir, en el marco del RIA, con carácter normativo, a qué concretos usos pueden o no, y, en su caso, con qué garantías, aplicarse soluciones de IA; como proveedora del sistema, en tanto en cuanto se implique en el diseño y desarrollo –aun cuando sea con la colaboración de terceros- de concretos sistemas de IA y, finalmente, como usuaria, por cuanto a ella corresponderá calibrar la oportunidad del uso de una concreta solución de IA preexistente para la consecución del concreto objetivo perseguido en cada caso.

Son ilustrativas al respecto las conclusiones del subgrupo de expertos en Inteligencia Artificial constituido en la OCDE (AIGO) en noviembre de 2018 para informar la Recomendación del Consejo sobre Inteligencia Artificial[44], que asumen las particularidades de los análisis del ciclo de vida de un sistema de IA respecto de los

---

43 Así lo pone de manifiesto J. ESTEVE PARDO (2023: 165-167), bajo las categorías de "Decisión, gestión y responsabilidad". El apartado 5.5 de la norma ISO-3100 "Gestión del Riesgo. Principios y Directrices", que establece el estándar internacional para la gestión del riesgo en cualquier ámbito, se expresa como sigue: "La selección de la opción más apropiada de tratamiento del riesgo implica obtener una compensación de los costes y los esfuerzos de implementación en función de las ventajas que se obtengan, teniendo en cuenta los requisitos legales, reglamentarios y de otro tipo, tales como la responsabilidad social y la protección del entorno natural. Las decisiones también se deberían tomar teniendo en cuentas los riesgos cuyo tratamiento no es justificable en el plano económico, por ejemplo, riesgos severos (consecuencias altamente negativas) pero raros (baja probabilidad".

44 La referencia está extraída de OCDE (2019).

referidos a otros sistemas informáticos. En este sentido, aquellos se caracterizan por cuatro fases: 1.- Fase de planificación y diseño (conforme a los objetivos deseados y los requisitos que de ello se deriven), datos (identificando los disponibles, afinando su calidad y su disponibilidad futura, de forma documentada) y modelización e interpretación (concentrada en los algoritmos a utilizar y, en su caso, su entrenamiento), que incluye variadas actividades cuyo orden puede variar entre distintos tipos de sistemas de IA. 2.- Fase de verificación y validación, que implica la ejecución, comprobación y ajuste del modelo en función de distintas consideraciones. 3.- Puesta en producción, incluyendo las evaluaciones de legalidad y organizativas, y 4.- Uso y seguimiento, incluyendo una comprobación continua de las recomendaciones e impactos, intencionados o no, en función de los objetivos marcados, con el fin de identificar problemas y solucionarlos a través de la reedición de las fases previas o, si fuera necesario, la retirada del sistema.

Quedan así identificados varios elementos críticos en los sistemas de IA, que se manifestarán con más o menos intensidad según el concreto sistema que en cada momento se someta a análisis de riesgos. Y entre ellos cobra especial interés el relativo al tratamiento de los datos personales que en su caso hubieran de utilizarse, disciplinado conforme al Reglamento (UE) 2016/679, por el que se aprueba el Reglamento General de Protección de Datos (en adelante, RGPD) bajo una lógica que ha contagiado la del análisis de riesgos en general[45]. Así lo podremos comprobar en los apartados siguientes, al describir primero el mecanismo de gestión de riesgos que el RIA exige para los sistemas de alto riesgo y, después, los mecanismos que el citado RGPD introduce para la mejor garantía de la protección de los datos personales que se manejen en el entrenamiento del sistema previo a su puesta en funcionamiento y, sucesivamente, para su aplicación al uso pretendido. Y decimos "mejor garantía", conscientes de que es ilusoria cualquier

[45] En aplicación del art. 32 RGPD, en combinación con la aplicación del principio de privacidad por diseño y por defecto (art. 25 RGPD), tal y como tendremos ocasión de destacar. Es muy significativo a este respecto que la AEPD (2020: 15) ponga de manifiesto que la garantía del cumplimiento de los requisitos del tratamiento de datos personales que exija una solución de IA pasa por asegurar que la solución misma, desde un punto de vista técnico, responda a una serie de parámetros comunes entre los que estaría la precisión, exactitud o medidas de error requeridos para el tratamiento o de la solución de IA en función de la métrica adecuada para medir su bondad, la consistencia entre los resultados del proceso de inferencia o la predictibilidad del algoritmo. Y a ello suma (Ibidem: 19) severas consideraciones sobre la responsabilidad de quien decida adoptar una solución de IA en el marco de un tratamiento, del que es responsable y, en cuanto tal, obligado a determinar sus "medios y fines", sin que pueda escudarse en las limitaciones de información o desconocimiento técnico para evadir tal responsabilidad.

intención de domeñar plenamente sistemas basados en el manejo de información masiva, sometidos a la aplicación de algoritmos que pueden adquirir vida propia dentro de arquitecturas –especialmente las neuronales- muy complejas. Se trata de calibrar, pues, más que de pretender eliminar, los riesgos en presencia.

Se hace, así, imprescindible insistir en la necesidad de que las Administraciones Públicas se muestren prudentes al valorar la oportunidad de implantar sistemas de IA. Y a tal efecto, la importación de algunos instrumentos diseñados en la órbita de la protección de datos puede ser especialmente útil, trascendiendo incluso su alcance más allá de lo que suponga la necesaria protección de los datos personales. Estamos pensando, en particular, en la Evaluación de Impacto que, como veremos, contempla el RGPD, cuya aplicación generalizada en relación con la valoración de las soluciones de IA defiende la Guía orientativa propuesta por el *European Law Institute* cuando se trata particularmente de aplicar estas soluciones a la toma de decisiones[46]. De este modo se podrá, además, facilitar un cauce de participación a los interesados, tal y como tendremos ocasión de ponderar al tratar de los mecanismos de transparencia.

La Guía referenciada forma parte del creciente elenco de instrumentos orientativos, elaborados a nivel europeo e interno, dirigidos a acompañar el proceso de decisión de implantación de un sistema de Inteligencia Artificial en la órbita de la actuación pública[47]. Instrumentos que pueden resultar útiles, a falta de guía oficial, a pesar de que la ENIA prometiera la aprobación de una a imagen y semejanza de la que publicó el Gobierno del Reino Unido en junio de 2019 –no muy precisa, conviene advertirlo-[48].

---

[46] Se trata de una guía elaborada en el seno del *European Law Intitute*, de la Universidad de Viena, bajo la coordinación de M. WIERZBOWSKI, y en la que ha participado la Profa. Velasco Rico [M. WIERZBOWSKI (2022)]. Muy significativamente, el documento pone de manifiesto que en más de un extremo se ha inspirado en las previsiones sobre la evaluación de impacto ambiental.

[47] MANZONI et al (2022) ofrecen, en el contexto del JRC, una propuesta de hoja de ruta para la adopción de proyectos de IA con 16 recomendaciones en 4 áreas de intervención, que son glosadas por TORRECILLA (2023:87-88). En el ámbito español es interesante el trabajo, de cita generalizada, de A. ZLOTNIK (2019), y las Guía que propone M. MERCHÁN (2020: 21-), así como las apreciaciones de R. MARTÍNEZ MARTÍNEZ (2019).

[48] La ENIA, en su página 59, cita, en efecto, como referente, la Guía aprobada en el Reino Unido (2019).

En todas estas guías es, de hecho, recurrente llamar la atención sobre la existencia o no de datos significativos y suficientes, en términos de cantidad y calidad, y sobre la necesaria valoración del "impacto" asociado a los errores potenciales, criterio que pasa por calibrar verdaderamente no solo la probabilidad de la materialización de los riesgos pronosticables, sino su propia densidad[49]. La disponibilidad de un equipo humano suficiente y preparado tampoco es cuestión menor, incluso en los casos –en absoluto infrecuentes- en los que la solución más eficiente pase por la contratación de las soluciones de IA.

El papel de la contratación pública en la selección y aplicación de soluciones de IA será, muy probablemente, la regla. Sea de soluciones ya contrastadas en el mercado, sea –las menos de las veces- para desarrollar una solución específica con el fin de dar cobertura a una determinada necesidad de la Administración Pública, en correlación con su doble papel de usuarias o proveedoras de soluciones de IA que, como anticipamos, contempla el RIA para las autoridades públicas. En este último caso, el procedimiento de asociación para la innovación puede jugar un papel fundamental, como pone de manifiesto la propia ENIA, por más que –conviene tenerlo presente desde ahora- muchos elementos necesarios para la IA, empezando por los algoritmos, están ya a disposición, incluso extensivamente bajo fórmulas de uso compartido, con lo que resulta inverosímil plantear el diseño de una solución de IA partiendo de 0. Sea como fuere, en el ámbito de la contratación pública, el requisito de acreditar la necesidad del contrato que impone el art.

---

[49] A este respecto tomamos como referencia a A. ZLOTNIK (2019), quien, al plantear la necesidad de calibrar la oportunidad de un sistema de IA frente a una solución "basada en programación tradicional, con código fuente totalmente transparente y predecible", pone en escena un doble criterio: a) si existen datos representativos y suficientes del fenómeno en estudio y b) la valoración de si el impacto asociado a los errores del sistema de IA es asumible –entendiendo que el impacto introduce criterios cualitativos para calibrar los errores posibles más allá de lo cuantitativo-. A ello suma, como un doble criterio de menor entidad, la disponibilidad de un equipo humano con conocimiento suficiente sobre IA y sobre el dominio del problema concreto que se pretende resolver y la disponibilidad de recursos computacionales suficientes. Cuestión previa es, entendemos, la verosimilitud misma de la opción de IA desde un punto de vista técnico en función de la actuación pública a desarrollar. M. MERCHÁN (2020: 21-), por su parte, ofrece una verdadera Guía para la toma de decisiones sobre la viabilidad del proyecto en la que apunta también a la disponibilidad de datos suficientes y de calidad, pero no expresamente al impacto de los posibles errores, si bien reclama como aspecto del sistema de IA sobre los que "reflexionar" que cumpla con las exigencias éticas, de privacidad, de responsabilidad y de explicabilidad y transparencia.

28 de la LCSP exige por sí mismo, en todo caso, un esfuerzo de maduración en sede de la entidad pública contratante que puede facilitar el tipo de análisis que ahora nos ocupa.

Esta llamada a la prudencia que defendemos y las concretas apreciaciones recién hechas, no suponen discutir el papel que desde las instancias europeas se viene reconociendo a las Administraciones Públicas como incentivadoras del desarrollo de soluciones de Inteligencia Artificial[50], y que la ENIA española, como anticipábamos, asume como propio. De hecho, ellas mismas están llamadas a facilitar proyectos acogidos a bancos de pruebas que, precisamente, sirvan para calibrar sus posibles riesgos antes de implantarlos en la práctica, en los términos 53 del RIA. España se presenta como pionera al respecto, habiendo cerrado el 31 de octubre pasado el plazo de preinscripción de las entidades y empresas interesadas[51], como anticipo a la aprobación del Real Decreto 817/2023, de 8 de noviembre, que establece un entorno controlado de pruebas para el ensayo del cumplimiento de la propuesta de RIA[52]. Se tratará, pues, de ensayar el cumplimiento de lo que no deja de ser, a día de hoy, un proyecto de norma, pero firmemente asentado en una lógica de la regulación de riesgos que le trasciende.

Muy significativo en relación con todo ello es que la Agencia Española de Supervisión de Inteligencia Artificial cuyo estatuto ha sido aprobado mediante Real Decreto 729/2023, de 22 de agosto, si bien no ha iniciado sus actividades, contemple en su art. 25.b).1º como una de las funciones del Departamento de Sistemas de Inteligencia Artificial orientados a las Administraciones Públicas integrado en la Subdirección de Informes e Infraestructuras de Pruebas, la de "emitir informes preceptivos respecto de aquellos pilotos que desplieguen sistemas de inteligencia artificial o espacios de prueba puestos en marcha desde los diferentes departamentos ministeriales, o desde cualquier

---

50 Ya desde la Comunicación de la Comisión Europea de 2018 "Inteligencia Artificial para Europa" [COM (2018) 237 final, de 25.4.2018]. Este y otros antecedentes del proyecto de RIA aparecen bien sintetizados en E. GÓMEZ et al. (2023: 728-730) y, más extensamente, en N. NIKOLINAKOS (2023: 23-306).

51 Se informa al respecto en la página web de la vicepresidencia primera: https://portal.mineco.gob.es/gl-es/comunicacion/Paxinas/231002_sandbox_ia.aspx

52 La CNMC ha dictado, con fecha 25 de julio de 2023, informe respecto del proyecto de real decreto que se puede consultar en https://www.cnmc.es/sites/default/files/4822555.pdf

entidad del Sector Público"[53]. Pero lo es más aún que la propia puesta en marcha de los sistemas de inteligencia artificial utilizados por las administraciones públicas quede supeditada al informe vinculante que emita el Departamento de certificación, instrucción y supervisión integrado en la Subdirección de Certificación, Evaluación de Tendencias, Coordinación y Formación en Inteligencia Artificial (sic), conforme establece el art. 26.a).6º del Real Decreto de referencia.

Habrá ocasión de volver sobre las coordenadas de esta Agencia, que está llamada a cumplir las funciones de supervisión que reclama la propuesta de RIA en su art. 59. Es, necesario, con todo, poner en este momento de manifiesto que resulta problemático reconocer a una Agencia estatal semejantes funciones de escrutinio previo a cualesquiera proyectos de IA desarrollados por cualesquiera Administraciones públicas, tanto por razones competenciales como por razones meramente prácticas.

La propuesta de RIA tiene, con todo, un impacto directo en los proyectos de uso de soluciones de IA que las Administraciones públicas puedan proyectar bajo las coordenadas que pasamos a describir.

B.- El impacto directo de la propuesta de RIA en los márgenes de utilización de sistemas de IA por Administraciones Públicas: usos prohibidos y de alto riesgo

La toma de decisiones por parte de las Administraciones en cuanto a la viabilidad de introducir soluciones de IA para el desempeño de sus tareas viene condicionada –para bien y para mal- por una premisa previa que impone, precisamente, la propuesta de RIA, toda vez que –como anticipamos- clasifica los sistemas de IA considerando algunos directamente prohibidos (art. 5) y otros de alto riesgo, para someterlos a un intenso escrutinio (art. 6)[54]. Siendo así que entre ellos se identifican -en uno y otro grupo- algunos que son de uso potencial, y en ocasiones exclusivo, por parte de las Administraciones Públicas: algunos cuya "finalidad prevista" –concepto clave en el RIA por cuanto parte

[53] El apartado 5º del mismo artículo identifica como otra función de este Departamento la de "realizar actividades e implementar herramientas de colaboración para apoyar el uso de esta tecnología entre las entidades del sector público".

[54] N. NIKOLINAKOS (2023: 374-445) ofrece una explicación pormenorizada de las variantes y condicionantes que se manejaron para determinar los distintos supuestos.

de una aproximación funcional en el análisis de riesgos- tiene que ver con una actuación pública[55].

A.- Entre los prohibidos, caracterizados como potenciales herramientas de manipulación, explotación y control social que se consideran intolerables por contradecir los valores de la Unión[56], están cualesquiera que se sirvan de técnicas subliminales o que se aprovechen de vulnerabilidades de un grupo específico de personas en función de su edad o discapacidad física o mental[57], pero además dos supuestos referidos específicamente a actuaciones de "autoridades públicas"[58].

En un primer caso, estrictamente en el ámbito penal, queda prohibido, salvo excepciones tasadas que se sometan a autorización judicial o de autoridad administrativa independiente en los términos que precise cada Estado miembro, el uso de sistemas de identificación biométrica remota "en tiempo real" en espacios de acceso público[59]. Por

---

55 Este concepto es clave, en efecto, en la lógica del RIA, en función de su aproximación basada en regular una tecnología como producto en función de unos riesgos que dependen de su concreto uso. De ahí que la "finalidad prevista" se defina en el apartado 112 del art. 3 del RIA como "el uso para el que un proveedor concibe un sistema de IA, incluido el contexto y las condiciones de uso concretas, según la información facilitada por el proveedor en las instrucciones de uso, los materiales y las declaraciones de promoción y venta, y la documentación técnica". Y que junto a ella se defina el "uso indebido razonablemente previsible" como "la utilización de un sistema de IA de un modo que no corresponde a su finalidad prevista, pero que puede derivarse de un comportamiento humano o una interacción con otros sistemas razonablemente previsible" (ap. 13 del mismo art. 3).

56 Según razona el cdo. 15 de la norma, "dichas prácticas son sumamente perjudiciales y deben estar prohibidas, pues van en contra de los valores de la Unión de respeto de la dignidad humana, libertad, igualdad, democracia y Estado de Derecho y de los derechos fundamentales que reconoce la UE, como el derecho a la no discriminación, la protección de datos y la privacidad, y los derechos del niño".

57 Aps. 1.a) y b).

58 El RIA utiliza en ocasiones la expresión "autoridad encargada de la aplicación de la ley" ceñida estrictamente, según se desprende de lo que precisa el apartado 40 del art. 3, a las "autoridades públicas" (expresión que no se define) con competencias relacionadas con las infracciones y sanciones penales, no administrativas.

59 Ap. 1.d) en conexión con los aps. 2 a 4. Entre los riesgos que estos sistemas generan la norma alude a su afección a la vida privada de una gran parte de la población, al efecto de provocar sensación de estar en vigilancia constante y disuadir indirectamente respecto del ejercicio del derecho de reunión y otros derechos fundamentales. Así se expresa el cdo. 18, mientras que los sucesivos 19 a 24 glosan los términos en los que cabe aplicar las excepciones tasadas, siendo especialmente relevante el tratamiento especial de los datos personales que, en caso de autorización excepcional, se obtengan, frente a los términos de lo previsto en el art. 10 de la Directiva (UE) 2016/680, 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

otra parte, se prohíben los sistemas de IA dirigidos a "evaluar o clasificar la fiabilidad de personas físicas durante un período determinado de tiempo atendiendo a su conducta social o a características personales o de su personalidad conocidas o predichas", de forma que la "clasificación social" resultante provoque un trato perjudicial o desfavorable hacia determinadas personas físicas o colectivos enteros en contextos sociales que no guarden relación con los contextos donde se generaron o recabaron los datos originalmente y/o un trato semejante que resulte desproporcionado a su comportamiento social o a la gravedad de este[60]. Prohibición esta última que puede tener importante impacto tanto en la órbita de actividades inspectoras –de personas físicas, conviene advertirlo- como de selección de beneficios de cualquier orden, si bien no determina la imposibilidad absoluta del manejo de sistemas de IA en tales ámbitos, siempre que no se incurra en las características que determinan la prohibición. Las previsiones del apartado 5 del Anexo III lo vienen, de hecho, a confirmar, como veremos inmediatamente, al considerar de alto riesgo sistemas con tales finalidades.

B.- Centrándonos, en efecto, en los sistemas que la propuesta de Ley europea de IA califica de alto riesgo, con la consecuencia de imponerles el cumplimiento de los requisitos que se desgranan en el Capítulo 2 del mismo Título III –y sobre los que habrá ocasión de reflexionar pormenorizadamente a lo largo de los sucesivos epígrafes-, debemos tener presente que la misma contempla dos escenarios en los dos apartados de su art. 6: de una parte, un amplio catálogo de supuestos en los que el sistema de IA debe ser, aisladamente o como componente de seguridad, objeto de una evaluación de conformidad conforme a la legislación de armonización que se especifica en el Anexo II (ap. 1); de otra, los sistemas que figuran en el Anexo III (ap. 2). En uno y otro caso, lo determinante para su consideración como de alto riesgo es su muy verosímil afección grave a tres valores tasados: la salud, la seguridad o los derechos fundamentales de las

ejecución de sanciones penales, y a la libre circulación de dichos datos y por la que se deroga la Decisión Marco 2008/977/JAI del Consejo, de modo que los datos así obtenidos no pueden ser objeto de utilización distinta.

60 Ap. 1.c).

personas[61], integrando entre estos un amplio catálogo que incluye el tan traído y llevado derecho a una buena administración, sobre el que volveremos[62].

Son los sistemas que aparecen en los ocho apartados de este Anexo III los que nos interesan particularmente[63], teniendo en cuenta que el mismo incorpora variados supuestos de sistemas de IA en los que, mayoritariamente, la "finalidad prevista" ha de corresponderse típicamente -cuando no exclusivamente- con una función de una Administración pública, en algunos casos por asimilación:

> 1.- Identificación biométrica remota de personas físicas –no en espacios abiertos, hay que entender, frente al supuesto prohibido en el art. 5.1-, por cuanto pueden generar resultados sesgados con efectos discriminatorios[64].
>
> 2.- Sistemas destinados a componentes de seguridad en la gestión y funcionamiento del tráfico rodado y en el suministro de agua, gas, calefacción y electricidad, en tanto en cuanto un fallo en los mismos puede poner en peligro la

---

61 Así lo razona el cdo. 27, en correlación con un planteamiento de mínima restricción del comercio internacional, en función de un criterio de proporcionalidad, lo que ha supuesto dejar fuera otros valores –como el medioambiente- que se habían reivindicado en este escenario, tal y como glosa N. NIKOLINAKOS (2023: 412-419). Lo determinante, en todo caso, es tanto la gravedad del posible perjuicio como la probabilidad de que se produzca, para determinar lo cual, en el caso particular de que se trate de sistemas de IA independiente de los contenidos en el Anexo III, se ha aplicado la metodología de análisis que se aplicará, a su vez, para su revisión, según se glosa en el cdo. 32.

62 En los términos del cdo. 28, "la magnitud de las consecuencias adversas de un sistema de IA para los derechos fundamentales protegidos por la Carta es particularmente pertinente cuando este es clasificado como de alto riesgo. Entre dichos derechos se incluyen el derecho a la dignidad humana, el respeto de la vida privada y familiar, la protección de datos de carácter personal, la libertad de expresión y de información, la libertad de reunión y de asociación, la no discriminación, la protección de los consumidores, los derechos de los trabajadores, los derechos de las personas discapacitadas, el derecho a la tutela judicial efectiva y a un juez imparcial, los derechos de la defensa y la presunción de inocencia, y el derecho a una buena administración". En relación con la afección de los derechos fundamentales por parte de la IA, desde una aproximación general, es ilustrativo el estudio de M. PRESNO LINERA (2023), así como, desde una perspectiva internacional, QUINTAVILLA, A. y TEMPERMAN, J. (Eds.) (2003), Artificial Intelligence and Human Rights, Oxford University Press.

63 Lo que no supone excluir radicalmente de la esfera de nuestros intereses los supuestos del Anexo II, en la medida en que los correspondientes productos queden involucrados en actividades de Administraciones Públicas. Téngase en cuenta, simplemente, el supuesto de los sistemas de diagnóstico y de apoyo a las decisiones humanas en el ámbito de la salud, que refiere el cdo. 28 de la norma.

64 Así lo argumenta el cdo. 33, como resultado de "imprecisiones técnicas", y con la consecuencia de que deban aplicárseles requisitos específicos referentes a las capacidades de registro y a la vigilancia humana.

vida y la salud de las personas a gran escala y alterar de forma apreciable el desarrollo habitual de las actividades económicas y sociales[65].

3.- Sistemas destinados a utilizarse para determinar el acceso o la asignación de personas físicas a centros de educación o formación profesional o para evaluar a los estudiantes de dichos centros o a los que pretendan acceder a los centros de educación, en función de los efectos discriminatorios que puedan suponer[66].

4.- Sistemas que, en relación con el empleo (en nuestro caso, público), se destinen a la contratación o selección de personas físicas, especialmente para anunciar puestos vacantes, clasificar y filtrar solicitudes o evaluar a candidatos en el transcurso de entrevistas o pruebas o se utilicen para tomar decisiones relativas a la promoción y resolución de relaciones contractuales "de índole laboral" –expresión que no debería impedir que se aplique igualmente a relaciones funcionariales-, a la asignación de tareas y al seguimiento y evaluación del rendimiento y la conducta de las personas en el marco de dichas relaciones, por cuanto pueden, particularmente, "perpetuar patrones históricos de discriminación"[67].

5.- Sistemas destinados –literalmente- "a ser utilizados por las autoridades públicas o en su nombre para evaluar la admisibilidad de las personas físicas para acceder a prestaciones y servicios de asistencia pública, así como para conceder, reducir, retirar o recuperar dichas prestaciones y servicios" –por cuanto suelen responder a situaciones de dependencia y vulnerabilidad y pueden afectar a derechos fundamentales-[68], así como los destinados "a utilizarse para el envío o el

---

65 Según razona el cdo. 34.

66 En el argumentario de la norma (cdo. 35), el potencial efecto discriminatorio puede afectar, en último término, a la capacidad de subsistencia de la persona, por condicionar su trayectoria formativa y profesional, y, en todo caso, "cuando no se diseñan y utilizan correctamente", pueden violar el derecho a la educación y a la formación y el derecho a no sufrir discriminación, "además de perpetuar patrones históricos de discriminación".

67 Esta parece ser, en efecto, la razón determinante para haber calificado de alto riesgo estas particulares finalidades de sistemas de IA, según se desprende del cdo. 36, que menciona también, si bien exclusivamente respecto de los sistemas de control de rendimiento y de comportamiento, la posible afección a la protección de los datos personales y a la privacidad. Siendo así, en el caso del empleo público y, en particular, en el caso de los funcionarios, que las razones últimas que laten en esta previsión derivan directamente del art. 23.2 CE.

68 El cdo. 37 asume que estos sistemas se utilizan "para decidir si las autoridades deben denegar, reducir, revocar o reclamar" ayudas y servicios de autoridades públicas, siendo así que "pueden afectar de un modo considerable a los medios de subsistencia de las personas y podrían infringir sus derechos fundamentales, como el derecho a la protección social, a la no discriminación, a la dignidad humana o a una tutela judicial efectiva". Su calificación como de

establecimiento de prioridades en el envío de servicios de primera intervención en situaciones de emergencia, por ejemplo bomberos y servicios de asistencia médica" –lo que podría entenderse extendido a los sistemas "de triaje" en la atención de enfermos en los propios servicios médicos-, toda vez que tales sistemas "adoptan decisiones en situaciones sumamente críticas para la vida y la salud de las personas y de sus bienes"[69].

6.- Sistemas relativos a la aplicación de la ley penal que se desgranan en el apartado 6, relacionados con la investigación o prueba de delitos[70], pero que bien podrían pensarse extrapolables, con las debidas adaptaciones, al ámbito del ejercicio de la potestad sancionadora, por cuanto tienen como fin último garantizar plenamente la protección de los derechos fundamentales en presencia[71].

7.- Sistemas de gestión de la migración, el asilo y el control fronterizo, que se concretan –con vistas a proteger a quienes se encuentran en situación de especial

---

alto riesgo, no debe suponer, sin embargo, que el RIA deba obstaculizar "el desarrollo y el uso de enfoques innovadores en la Administración pública, que se beneficiarían de una mayor utilización de sistemas de IA conformes y seguros, siempre y cuando dichos sistemas no conlleven un alto riesgo para las personas jurídicas y físicas".

[69] El mismo cdo. 37, *in fine*.

[70] Entre ellos se mencionan algunos particularmente relevantes a nuestros efectos, como los sistemas destinados a llevar a cabo evaluaciones de riesgos individuales de personas físicas con el objetivo de determinar el riesgo de que cometan infracciones penales o reincidan en su comisión, así como el riesgo para las potenciales víctimas de delitos –como ha sido el caso, de hecho, del sistema español VioGen- (subap. a); los destinados a utilizarse para la evaluación de la fiabilidad de las pruebas durante la investigación o el enjuiciamiento de infracciones penales (subap. d); los destinados a utilizarse para predecir la frecuencia o reiteración de una infracción penal real o potencial con base en la elaboración de perfiles de personas físicas, o para la elaboración de tales perfiles durante la detección, investigación o enjuiciamiento de infracciones penales, de conformidad con lo dispuesto en la Directiva (UE) 2016/680, o en la evaluación de rasgos y características de la personalidad o conductas delictivas pasadas de personas físicas o grupos (subaps. e y f); o los sistemas destinados a llevar a cabo análisis sobre infracciones penales en relación con personas físicas que permitan examinar grandes conjuntos de datos complejos vinculados y no vinculados, disponibles en diferentes fuentes o formatos, para detectar modelos desconocidos o descubrir relaciones ocultas en los datos (subap. g).

[71] El cdo. 38 desglosa los concretos riesgos que justifican esta clasificación, puntualizando las concretas garantías que pueden mitigarlos, siendo así que en más de un caso –como las referidas a la transparencia y a la explicabilidad y a la suficiente documentación, en el caso de los riesgos para los derechos de defensa y la presunción de inocencia- son perfectamente trasladables al ámbito de las infracciones administrativas. Conclusión que no queda descartada por el dato de que el mismo cdo. concluya negando que los sistemas de IA destinados específicamente a que las autoridades fiscales y aduaneras los utilicen en procesos administrativos deban considerarse dentro de la categoría que el artículo de referencia establece. Que el Reglamento no los considere de alto riesgo no impide que desde la normativa administrativa o, en todo caso, por parte de la autoridad administrativa competente, en su caso por vía contractual, se incorporen algunas garantías relevantes para conjurar riesgos equivalentes.

vulnerabilidad y dependencia[72]- en los que pretendan detectar el estado emocional de las personas físicas, en los que se destinen a utilizarse para evaluar el riesgo que plantee una persona física que pretenda entrar o haya entrado ya en el territorio de un Estado miembro respecto de cuestiones como la seguridad, la salud o la inmigración ilegal, en los destinados a utilizarse para la verificación de la autenticidad de la documentación de los migrantes y en los destinados a ayudar a examinar las solicitudes de asilo, visado y permisos de residencia y las correspondientes reclamaciones.

8.- Y, finalmente, sistemas "*destinados a ayudar a una autoridad judicial en la investigación e interpretación de hechos y de la ley, así como en la aplicación de la ley a un conjunto concreto de hechos*".

Este último supuesto, aunque referido estrictamente a las autoridades judiciales, resulta realmente revelador, ya que asume la posibilidad de utilizar sistemas de IA en la aplicación de la ley, aunque –significativamente- asumiendo su mero carácter auxiliar y no sustitutivo. Y como tal, resultaría perfectamente extrapolable, de forma adaptada, al ámbito del ejercicio de potestades administrativas de alcance decisor, con el mismo fin –confesado en el cdo. 40 del RIA- de evitar los posibles "sesgos, errores y opacidades" que pudieran producirse, aun cuando en el ámbito de la actuación administrativa, no se presentan en plenitud los riesgos de afección a "la democracia, el Estado de Derecho, las libertades individuales y el derecho a la tutela judicial efectiva y a un juez imparcial" que vienen a justificar la clasificación como de alto riesgo.

Es de advertir, en todo caso, que en el marco de la propuesta de RIA, cualquier extensión de sus previsiones no puede producirse por vía interpretativa. Ni tampoco parece que pueda producirse por la vía de la modificación del Anexo III que el art. 7 deja en manos de la Comisión, si bien bajo parámetros muy estrictos, el primero de los cuales es, precisamente, la imposibilidad de ampliar los ámbitos prefijados en los apartados 1 a 8 (art. 7.1.a), a lo que se suma la limitación de los criterios de justificación a lo que tenga que ver con la salud y seguridad de las personas y la protección de los derechos fundamentales. En la lógica de la propuesta de RIA, la armonización de reglas que

[72] Según se desglosa en el cdo. 39, que incide en que los correspondientes sistemas deben ser "precisos, no discriminatorios y transparentes".

ordenan la puesta en el mercado o en servicio de un sistema de IA, entendido como producto, solo excepcionalmente puede justificar una prohibición o, aún con menor intensidad, una calificación como de alto riesgo cuya consecuencia es la imposición de una serie de severos requisitos que, con carácter general, quedan sometidos a un sistema de autocontrol de conformidad, que solo en el caso del apartado 1 –identificación biométrica y categorización de personas físicas- puede llegar a incorporar una evaluación de conformidad emitida por un organismo notificado previa a su puesta en servicio (art. 43). Y ello sin perjuicio del deber de registro previo de los sistemas de alto riesgo contemplados en el art. 6.2 en la base de datos de la UE creada al efecto (arts. 51 y 60) y del sometimiento, en su caso, al régimen sancionatorio cuya concreción remite el art. 71 del RIA a lo que concreten los Estados miembros, a los que les compete también decidir si podrán imponerse o no multas a las autoridades y organismos públicos internos.

Más allá de su ámbito imperativo, el RIA anima al alineamiento con sus valores propiciando la suscripción de códigos de conducta de adhesión voluntaria (art. 69). Bajo lógica equivalente operarían los sellos de calidad que pone en escena el art. 23 de la Ley 15/2022 y en cuyo diseño está avanzando el Ministerio de Asuntos Económicos y Transformación Digital una vez ha adjudicado, el pasado octubre, la correspondiente licitación[73]. El alcance y proyección de unos y otros es general, si bien las Administraciones públicas tienen instrumentos para trascender de las coordenadas imperativas o voluntarias que pone en escena esta propuesta de Reglamento. Podrá hacerlo al regular el uso de IA en el ámbito administrativo, yendo más allá de las previsiones indiciarias de la misma Ley 15/2022 o del 41 LRPAC –sobre cuyas limitaciones habrá ocasión de reflexionar-. Y también en cuanto que proveedora de un concreto sistema de IA, o, aún como usuaria, a través de la correspondiente licitación del sistema, introduciendo requisitos iguales o equivalentes a los que el RIA exige para los sistemas considerados de alto riesgo[74].

---

[73] En el BOE de 21 de octubre de 2023 se publicó la adjudicación a la UTE constituida entre Deloitte y Odiseia del contrato de servicios para el desarrollo de un sello de calidad de IA conforme a la licitación cuyo anuncio se publicó el 9 de enero de este mismo año. Los adjudicatarios disponen de 30 meses para la ejecución del contrato desde la formalización.

[74] Es muy expresivo a este respecto el planteamiento de la Propuesta de cláusulas contractuales tipo para la contratación de Inteligencia Artificial que han sido auspiciadas por la Comisión Europea, distinguiendo según se trate o no de sistemas considerados de alto riesgo conforme a la propuesta de RIA. El pasado octubre se publicó una segunda versión de las mismas: PUBLIC BUYERS COMMUNITY (2023).

Como consecuencia de todo lo anterior, cabe concluir que en la toma de decisión por una Administración Pública de si utilizar o no un sistema de IA, la decisión le viene dada, en negativo, para los sistemas prohibidos por el art. 5 del RIA, y le viene condicionada, en cuanto al cómo, en el caso de los sistemas de IA calificados de alto riesgo por el propio RIA. Un cómo que, en la medida en que los somete a requisitos severos dirigidos a revestir su uso de garantías, sirve por sí mismo a los fines de aseguramiento de la más plena conformidad de la actuación pública al ordenamiento jurídico, trascendiendo incluso los concretos criterios que preocupan a la Unión Europea cuando del uso de IA se trata. En lo que se refiere a cualesquiera otros sistemas de IA que una Administración Pública se plantee desarrollar o utilizar, las pautas que le ofrece el RIA en su delimitación de garantías respecto de sistemas de alto riesgo le han de servir de guía muy valiosa para, aun aplicando criterios de ponderación y proporcionalidad, identificar los ámbitos de atención ineludibles[75]. Empezando por la determinación de un sistema de gestión de riesgos como el que el art. 9 del RIA exige para los sistemas de alto riesgo, cuyo planteamiento está sustanciado –como anticipamos y pasamos a desarrollar- en el principio de ciclo de vida.

2.- La introducción de un mecanismo continuado de gestión de riesgos como clave de bóveda del modelo regulatorio conforme al principio del ciclo de vida: especial énfasis en los requisitos de trazabilidad y seguridad en el contexto del Esquema Nacional de Seguridad

Teniendo en cuenta que la perspectiva regulatoria que asume la Unión Europea respecto de la IA se basa en un afán de domeñar los riesgos potenciales a sus distintas manifestaciones y aplicaciones, no es de extrañar que el primer requisito que el RIA contemple para los sistemas de IA de alto riesgo se refiera a la obligada implantación, documentación y mantenimiento de un sistema de gestión de riesgos. Este sistema funcionará, además, tal y como se hace expreso en el art. 8.2 de la norma, como un elemento a tener en cuenta a la hora de verificar el cumplimiento de los demás requisitos,

[75] WIERZBOWSKI, M. (Coord.) (2022) asume que los sistemas utilizados por las Administraciones Públicas para la toma de decisiones pueden o no ser de alto riesgo, a criterio de cada ordenamiento, con la consecuencia de someterlos a requisitos más o menos severos en cuanto a la evaluación de su implantación y continuación.

que en último confluyen en él, y se integrará a su vez dentro del sistema de calidad al que posteriormente haremos referencia.

En estos términos, sobre la base del criterio del ciclo de vida, ya explicado, el sistema de gestión de riesgos se presenta como la última etapa del proceso de diseño del sistema de IA, que se va construyendo a través de una lógica de "prueba y error" en la constatación y corrección o mitigación de los riesgos inherentes al sistema, analizados en función de su finalidad prevista, hasta su definitiva puesta en servicio o entrada en el mercado, siendo así que el sistema de gestión de riesgos permite mantener la misma lógica en la fase operativa. Se trata, pues, de un proceso iterativo continuo a llevar a cabo durante todo el ciclo de vida del sistema con las consiguientes actualizaciones sistemáticas periódicas (art. 9.2). Como técnica típica que es de la regulación de riesgos, con él se trata, en definitiva, de asegurar una identificación y análisis de los riesgos del sistema que no se agote en el momento de diseño y desarrollo del mismo, sino que se mantenga de forma continuada sobre la base de los resultados constatados en su aplicación práctica en los términos del art. 61, siempre incluyendo la previsión de medidas a adoptar conforme a lo previsto en los apartados 3 y 4, en conexión con el 2.d), del mismo art. 9.

En la fase de gestión del riesgo, lo determinante serán, pues, las medidas a adoptar. Para determinar cuáles sean las más adecuadas los sistemas se someterán a las oportunas pruebas durante la fase de desarrollo, conforme a los parámetros y umbrales de probabilidades previamente definidos en atención a su finalidad, en los términos de los apartados 5 a 7. Para su fijación, el apartado 3 prescribe que se tomarán en consideración los efectos de la aplicación combinada del resto de requisitos que se establecen en el RIA para los sistemas de alto riesgo, siempre teniendo en cuenta el "estado actual de la técnica generalmente reconocido", procurando –en los términos del apartado 4- eliminar o reducir los riesgos en la medida de lo posible mediante un diseño y desarrollo adecuado, a salvo los riesgos residuales, de los que se informará al usuario; implantar medidas de mitigación y control de los riesgos que no puedan eliminarse y, en todo caso, proporcionar a los usuarios la información oportuna conforme al art. 13, particularmente respecto de los riesgos inherentes al uso conforme a la finalidad prevista o según un uso indebido razonablemente previsible, y, en su caso, la oportuna formación.

Esta última advertencia conecta con la elocuente precisión final del propio apartado 4, que advierte que cuando se eliminen o reduzcan los riesgos asociados a la utilización del sistema de IA de alto riesgo, "*se tendrán en la debida consideración los conocimientos técnicos, la experiencia, la educación y la formación que se espera que posea el usuario, así como el entorno en el que está previsto que se utilice el sistema*". Esta consideración pone en escena la importancia de la formación del personal al servicio de las Administraciones que haya de involucrarse en el uso de sistemas de Inteligencia Artificial, particularmente en conexión con la cuestión de la vigilancia humana, que posteriormente abordaremos de forma destacada.

En los planteamientos del RIA en relación con este elemento nuclear de su regulación de los sistemas de IA, se hace evidente que la propia UE es consciente de que no es posible reducir a 0 los riesgos. Se trata, pues, de calibrarlos –en función de su gravedad y probabilidad- y de ponderarlos –en función de su admisibilidad respecto de los bienes y valores sobre los que inciden- para, una vez han sido asumidos conforme a parámetros de proporcionalidad en sentido propio –bajo la lógica coste-beneficio-, plantear las medidas de mitigación más oportunas. El RIA no deja de ser consciente, en definitiva, de que en el "estado actual de la técnica generalmente reconocido" –salvaguarda que con todo sentido embrida todas las obligaciones derivadas de la gestión de riesgos-, muchas de las garantías que se postulan tienen una virtualidad relativa, aun cuando su introducción esté al servicio de hacer "confiable" la IA.

Este relativismo no ha de suponer renuncia, en todo caso. De modo que la conveniencia de contar con un sistema de gestión de riesgos es predicable de cualquier sistema de IA que empleen las Administraciones Públicas, sea o no considerado de alto riesgo conforme al RIA, supuesto en el que resulta obligado conforme a todos los parámetros señalados antes de su puesta en servicio. La existencia de un tal sistema y su debida aplicación y seguimiento, es una garantía en sí misma para la implantación de soluciones de IA en cualquier aspecto de la actuación administrativa y, en cuanto tal, debería exigirse con carácter previo, siempre y en todo caso, con independencia de que su calado será distinto en función del sistema y la finalidad prevista, por cuanto los riesgos a gestionar serán de mayor o menor intensidad y/o manejabilidad. Y ello siempre partiendo de que, si el sistema ya está operativo, se habrá calibrado que los riesgos a gestionar son, en último

término, asumibles, por cuanto, en caso contrario, el análisis previo de la viabilidad y conveniencia de la incorporación de una solución de IA habría descartado su adopción, según reflexionamos en el epígrafe anterior.

Tal y como advertíamos, el sistema de gestión de riesgos de un sistema de IA alto riesgo formará parte del sistema de gestión de la calidad que el art. 17 del RIA establece como obligación de sus proveedores con el fin de garantizar el cumplimiento del Reglamento. El apartado 1.g) de este artículo se refiere, en efecto, al sistema de gestión de riesgos, entre otros muchos extremos que deben quedar documentados de forma sistemática y ordenada mediante políticas, procedimientos e instrucciones escritas. Se trata, con todo, de instrumentos internos, por más que su contenido pueda coincidir en parte con el de la documentación técnica que los arts. 11 y 18, en conexión con el anexo IV, exigen se prepare antes de la introducción al mercado o puesta en servicio del sistema de alto riesgo, con el fin de demostrar que cumple todos los requisitos, facilitando la oportuna evaluación de las autoridades competentes y los organismos notificados. Documento aparte es la declaración UE de conformidad que el proveedor deberá redactar para cada sistema, manteniéndola a disposición de las autoridades nacionales competentes durante un período de diez años, según le exige el art. 48 RIA.

El RIA regula los sistemas de IA como un producto. No lo olvidemos. Coherentemente con ello, el proyecto de norma establece dos reglas suplementarias con claro impacto en las garantías de los sistemas, que, en el ámbito de las Administraciones Públicas, se reconduce al cumplimiento del Esquema Nacional de Seguridad, él mismo basado en el análisis y gestión de riesgos.

A los efectos de poder permitir controlar el funcionamiento del sistema en el caso de que presente un riesgo sobrevenido en los términos del art. 65 y, en todo caso, para facilitar el seguimiento posterior a la comercialización en los términos del 61, el art. 12 exige que el diseño y desarrollo de los sistemas contemple expresamente la posibilidad de registrar automáticamente eventos durante el funcionamiento, facilitando "archivos de registros", permitiendo así una trazabilidad del funcionamiento del sistema durante el ciclo de vida[76].

[76] Para los sistemas de identificación biométrica se introducen exigencias más precisas en el apartado 4.

Esta obligación ha de jugar un importante papel en relación con los sistemas utilizados por Administraciones Públicas, tal y como tendremos ocasión de razonar al tratar de la transparencia, siendo a tal fin determinante el cumplimiento del Esquema Nacional de Seguridad que contempla el art. 156.2 LRJSP, tal y como se desprende del art. 1.2 del Real Decreto 311/2022, de 21 de mayo, que lo regula[77].

El Esquema se alinea también, por concepto, en la fase operativa, con las exigencias del RIA de que los sistemas de IA de alto riesgo se diseñen y desarrollen asegurando un nivel adecuado de precisión, solidez y ciberseguridad durante todo el ciclo de vida, siempre en función de la finalidad prevista, pudiendo aplicar al respecto soluciones de redundancia técnica como copias de seguridad o planes de prevención contra fallos, e introducir soluciones técnicas para subsanar las vulnerabilidades específicas de la IA, como las que pretendan manipular los datos de entrenamiento ("contaminación de datos") o los datos de entrada con la finalidad de que el sistema cometa un error ("datos adversarios").

No es en absoluto casual la importancia que da el RIA a las medidas de protección contra los ataques consistentes en la manipulación de los datos manejados por el sistema. En el análisis de riesgos de un sistema de IA, y particularmente en la fijación de las medidas de gestión de los detectados, un elemento de especial atención serán, sin duda, los datos manejados en la fase de entrenamiento, si corresponde, y en todo caso en la fase de aplicación. Cuando se trate de datos personales, la densa normativa que los protege, partiendo del RGPD, proyecta sobre la regulación de los sistemas de IA una serie de exigencias, que, estando como están igualmente basadas en la lógica de la regulación de riesgos, se solapan, particularmente desde el punto de vista procedimental, con los que el propio RIA establece. Se reafirma así, y en algún caso se amplifica, el catálogo de instrumentos de garantía a tener en cuenta para hacer viable el uso de sistemas de IA por Administraciones Públicas.

---

[77] En sus términos literales, "el ENS está constituido por los principios básicos y requisitos mínimos necesarios para una protección adecuada de la información tratada y los servicios prestados por las entidades de su ámbito de aplicación, con objeto de asegurar el acceso, la confidencialidad, la integridad, la trazabilidad, la autenticidad, la disponibilidad y la conservación de los datos, la información y los servicios utilizados por medios electrónicos que gestionen en el ejercicio de sus competencias". En su art. 14 reclama expresamente la implantación de un sistema de análisis y gestión de riesgos por cada organización.

3.- Mecanismos de garantía en la gestión de los datos frente a los riesgos de vulneración de la privacidad y la igualdad

Los datos son la materia prima de la Inteligencia Artificial. En los datos, más aun incluso que en los algoritmos, se juega cada vez más la fiabilidad de los sistemas de IA, particularmente cuando los sistemas se basan en algoritmos entrenados con datos.

Solo si se dispone de datos en la cantidad y con la calidad precisas para la concreta finalidad de cada sistema será viable una solución de IA, tanto desde un punto de vista técnico, como desde la perspectiva de su imprescindible alineamiento con los requisitos de protección de los derechos fundamentales en presencia. Y más concretamente, el derecho a la protección de los datos personales (art. 18.4 CE) –cuando se disponga de ellos para el entrenamiento del modelo o para su utilización, particularmente en el caso de decisiones automatizadas- y el derecho a la igualdad y a no sufrir discriminación (art. 14 CE) -puesto en riesgo cuando los sistemas de IA generan sesgos, deliberados o no-[78]. A la minimización de estos apela expresamente, como nos consta, el art. 23.1 de la Ley 15/2022, incluyendo la precisión de que los mecanismos que se apliquen al efecto habrán de incluir su diseño y datos de entrenamiento (de los algoritmos, hay que entender, siempre que sea pertinente), abordando su potencial impacto discriminatorio, a cuyo fin se promoverá la realización de evaluaciones de impacto.

Esta última precisión del artículo que utilizamos como cabecera no es en absoluto casual. La minimización de los riesgos que pueda implicar la implantación de un sistema de IA por parte de una Administración Pública encuentra, como venimos insistiendo, una fundamental piedra de toque en las cautelas relativas a la recopilación y tratamiento de

[78] A. ZLOTNIK (2019: 27) ofrece ejemplos muy ilustrativos al respecto. Particularmente respecto de las exigencias de calidad, E. GÓMEZ et al (2023: 731) aclaran que se trata de impedir riesgos como el de la impredicibilidad, que se deriva de la falta de representatividad de los datos que se le aportan al sistema de IA basado en técnicas de aprendizaje, o de la falta de generalización o sobreajuste de los mismos. La cantidad, por su parte, no es un fin en sí mismo, ya que como pone de manifiesto el Informe de la AEPD (2016), titulado "10 malentendidos relacionados con la anonimización", sobre el que volveremos, un aumento sustancial de la cantidad de datos en el conjunto de datos de entrenamiento puede derivar en vulnerabilidades respecto a la privacidad, por estar el modelo ajustado a datos fácilmente reidentificables, o en sesgos. Esta llamada a la contención en la cantidad de los datos a manejar es coherente con el principio de minimización que se enuncia en el art. 5.1.c RGPD, sobre cuyo contenido –que tiene un alcance absoluto o relativo, en función del "descreme" de los datos- reflexiona AEPD (2020: 38-41), con referencia expresa a las múltiples técnicas de minimización de datos que se aplican específicamente para aplicaciones de IA, en particular de *Machine Learning*.

los datos que se manejen, particularmente cuando se trata de datos personales. Se da así entrada a la regulación específica de estos datos, que maneja una serie de técnicas bien decantadas para la gestión de los riesgos inherentes a su tratamiento también basada en la lógica del ciclo de vida y en la que las evaluaciones de impacto están llamadas a jugar un importante papel.

Conviene, con todo, enfatizar una vez más que el riesgo 0 no existe. Es más, como la propia expresión de la Ley 15/2022 pone de manifiesto, en coherencia con el planteamiento del RGPD y de la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales (en adelante, LOPDGDD), la aspiración se agota en una "minimización" de los riesgos. No otro es tampoco el planteamiento de la propuesta de RIA, como nos consta se extrae de su art. 9[79], por más que el art. 10 introduzca criterios de calidad de los datos en relación con los sistemas de IA de alto riesgo, tanto para los que requieran entrenamiento de modelos con datos, como para los que no.

Para los primeros se introducen criterios de calidad exhaustivos en relación con los conjuntos de datos de entrenamiento, validación y prueba, exigiendo que se sometan a "prácticas adecuadas de gobernanza y gestión de datos"[80]; que cumplan con los criterios de pertinencia, representatividad, exactitud y completud, exigiendo que tengan las propiedades estadísticas adecuadas, también en lo que respecta a las personas o los grupos de personas en relación con los que se pretenda utilizar el sistema, cuando proceda, y, finalmente, que tengan en cuenta, en la medida necesaria en función de su finalidad

[79] Tengamos presente que conforme a su apartado 2, las medidas de gestión de riesgos darán la debida consideración a los efectos y posibles interacciones derivados de la aplicación combinada de los requisitos estipulados en el mismo Capítulo 2.

[80] Que según precisa el apartado 2 se centrarán en particular en la elección de un diseño adecuado; en la recopilación de datos; las operaciones de tratamiento oportunas para la preparación de los datos, como la anotación, el etiquetado, la depuración, el enriquecimiento y la agregación; la formulación de los supuestos pertinentes, fundamentalmente en lo que respecta a la información que, ateniéndose a ellos, los datos miden y representan; la evaluación previa de la disponibilidad, la cantidad y la adecuación de los conjuntos de datos necesarios; la detección de posibles lagunas o deficiencias en los datos y la forma de subsanarlas y, particularmente el examen atendiendo a posibles sesgos.

prevista, las características o elementos particulares del contexto geográfico, conductual o funcional específico en el que se pretende utilizar el sistema[81].

Para los sistemas que no requieran el entrenamiento de modelos, el mismo art. 10, en su apartado 6, requiere –de forma más laxa- que se empleen prácticas adecuadas de gobernanza y gestión de datos con vistas a garantizar que dichos sistemas cumplen lo dispuesto en el apartado 2.

Resulta así que, requiera o no el sistema de IA de alto riesgo entrenamiento de datos, la clave está, en el marco del RIA, en aplicar las prácticas de gobernanza y gestión de datos que resulten en cada caso oportunas para asegurar que aquellos son, en cantidad y, sobre todo, en calidad, adecuados a los fines del sistema y que no generan riesgos inasumibles para los derechos fundamentales en presencia, particularmente el derecho a la privacidad y el derecho a la igualdad y no discriminación[82]. Bajo estas coordenadas, es indudable que el manejo de sistemas de IA por las Administraciones Públicas no puede renunciar a incorporar estas cautelas, aun cuando no se trate de sistemas de alto riesgo, lo que se ve facilitado por las exigencias de carácter horizontal que, en particular respecto del primer ámbito, impone la normativa de protección de datos toda vez que la descripción de tareas que el art. 10 RIA pone en escena son identificables sin duda, cuando tienen por objeto datos personales, como tratamiento de los mismos en los términos del art. 4.2 RGPD[83].

A.- La normativa de protección de datos personales como parámetro de garantía para la utilización de sistemas de IA por Administraciones Públicas: auditorías, certificaciones y evaluaciones en escenarios complejos

---

81 Según exige el apartado 4, siendo así que se presumirá que cumplen este requisito los que hayan sido entrenados y probados con datos relativos al entorno geográfico, conductual y funcional específico en el que esté previsto su uso, en los términos del art. 61.2 del RIA.

82 En esta línea razona, de hecho, el cdo. 44 del RIA.

83 Que lo identifica con "cualquier operación o conjunto de operaciones realizadas sobre datos personales o conjuntos de datos personales, ya sea por procedimientos automatizados o no, como la recogida, registro, organización, estructuración, conservación, adaptación o modificación, extracción, consulta, utilización, comunicación por transmisión, difusión o cualquier otra forma de habilitación de acceso, cotejo o interconexión, limitación, supresión o destrucción". La toma de decisiones sobre una persona física a partir de datos personales (art. 22), que puede implicar en su caso la elaboración de perfiles (art. 4.4), se consideran expresamente tratamiento en los cdos. 24 y 72 RGPD, según recuerda AEPD (2020:10).

Efectivamente, por lo que hace a la garantía del derecho a la privacidad en relación con la gestión de datos inherente a los sistemas de IA, es indudable que es necesario tener presente la regulación dirigida, en particular, a la protección de datos personales, entendidos estos, en los términos del art. 4.1) RGPD[84], como "toda información sobre una persona física identificada o identificable"[85].

Tal y como ha puesto de manifiesto la AEPD en su documento "Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción", de febrero de 2020, los sistemas de IA pueden implicar tratamientos de datos personales –aunque no sea necesariamente así- en dos tipos de escenarios: por parte del propio sistema, en la medida en que utilice datos personales, originalmente en su entrenamiento y/o en su explotación; o, en el caso de decisiones automatizadas de las que regula el art. 22 RGPD, en la medida en que estas puedan afectar a personas, sea porque ofrezcan predicciones sobre el comportamiento de un sujeto, sea porque evalúen su estado actual, sea porque decidan ejecutar determinadas acciones respecto de él, pudiendo –en definitiva- servir de apoyo para la toma de decisiones por un ser humano, o adoptar por sí mismos decisiones. Todo ello en el bien entendido de que "el componente IA no va a estar aislado, sino que va a formar parte de un tratamiento más amplio"[86], lo que viene a complicar el establecimiento de responsabilidades y garantías.

En unos casos, los menos frecuentes, la solución de IA se desarrollará específicamente para el tratamiento en que se integre; en la mayoría de ellos, sin embargo, el componente de IA será desarrollado por terceros distintos del responsable del tratamiento de datos, muy típicamente sin intención de exclusividad. Lo que ocurre es que, conforme al criterio

---

84 Es significativo a este respecto que el Reglamento (UE) 2021/694, por el que se establece el Programa Europa Digital, en su art. 5.1 *in fine*, al referirse a Objetivo específico 2, relativo a la Inteligencia Artificial, para cuyo cumplimiento la Unión aportará contribución financiera para la consecución de los objetivos operativos que se establecen en el mismo apartado 1, hace expreso –en términos que no pueden ser sino, un mero recordatorio enfático- que "*las soluciones basadas en la IA y los datos que se faciliten deberán respetar el principio de privacidad y seguridad desde el diseño y deberán respetar plenamente la normativa sobre protección de datos*".

85 Precisando que "se considerará persona física identificable toda persona cuya identidad pueda determinarse, directa o indirectamente, en particular mediante un identificador, como por ejemplo un nombre, un número de identificación, datos de localización, un identificador en línea o uno o varios elementos propios de la identidad física, fisiológica, genética, psíquica, económica, cultural o social de dicha persona".

86 Según argumenta en AEPD (2020: 6).

del ciclo de vida, y por la naturaleza misma de las soluciones de IA, particularmente las basadas en entrenamiento de datos, el que haya desarrollado la solución ha incurrido en tratamientos de datos que pueden condicionar severamente el resultado del tratamiento de datos que lleve a efecto el usuario de la solución de IA, a lo que se suma que en la fase de explotación la entidad que requiere tal solución –por lo que nos interesa, una Administración Pública- puede disfrutar de ella explotándola por sí misma, aunque haya sido desarrollada por otro, o puede contratar al desarrollador o incluso a un tercero para su explotación. En el primer caso sería ella misma la responsable del tratamiento, en el segundo el desarrollador o el tercero serían los responsables o los encargados del mismo, según actúen independientemente o siguiendo instrucciones de la Administración[87].

Teniendo estos parámetros claros, aun dentro de su complejidad -que redunda en la necesidad de que la Administración actuante solo tome una decisión de acudir a una solución de IA sobre la base de un análisis profundo y honesto-, quienquiera que en cada estadio o función sea responsable del tratamiento de datos personales debe asegurarse de que se cumplen todos los parámetros de legalidad que impone el RGPD. Solo así se podrá asegurar la plena validez de los actos que traigan causa de los sistemas de IA aplicados, muy en particular si ellos mismos apoyan o determinan la decisión.

A tal fin, al responsable le corresponde adoptar las medidas técnicas y organizativas adecuadas (art. 24) y asegurar la protección de datos desde el diseño y por defecto bajo la lógica del ciclo de vida (art. 25), teniendo en cuenta en particular –tanto el responsable como el encargado del tratamiento, en los términos del art. 28 LOPD- los supuestos de potencial mayor riesgo que el mismo identifica. Junto a ello, con carácter general, a él corresponde asegurar el debido cumplimiento de los principios enunciados en el art. 5,

[87] En los términos de AEPD (2020: 12-14), "a la hora de realizar un análisis del tratamiento, el componente de IA y el resto de los elementos que conforman el tratamiento se han de estudiar como un todo". En todo caso, en el ciclo de vida de una solución de IA, se pueden encontrar tratamientos de datos personales no solo en las etapas de entrenamiento, sino también de validación, despliegue (en el supuesto usual de que la solución de IA sea un componente que se ofrece a terceros), explotación (en sus fases de inferencia, decisión o evolución) y retirada. Ibidem (17-19) ofrece un cuadro exhaustivo, muy ilustrativo, de los distintos supuestos a contemplar en función de las distintas etapas del ciclo de vida del sistema de IA, si bien advierte que contempla los supuestos más comunes, que pueden complicarse en los supuestos de redes cooperativas, tipo *blockchain*, o en el no poco común de intervinientes que puedan recoger datos en el marco del Big Data para ofrecérselos a los desarrolladores. Un catálogo muy extenso y muy complejo de supuestos, en definitiva.

como en particular, además de la mencionada transparencia (ap. a), el de limitación de la finalidad (ap. b), minimización (ap. c) y el de limitación del plazo de conservación (ap. e). Y ello en el bien entendido de que estos principios, como las medidas de seguridad que igualmente deben introducirse en función del análisis de riesgos en su momento establecido conforme a la lógica del ciclo de vida (art. 32) y las propias exigencias del art. 22, pueden ser moduladas a través de medidas legislativas adoptadas por Derecho de la Unión o interno en los supuestos y con los condicionantes establecidos en el art. 23.

Las mencionadas medidas técnicas y organizativas de seguridad deberán adaptarse a cada concreto tratamiento en función de sus fines, medios y circunstancias, siempre sobre la base de un análisis veraz de los riesgos y amenazas. Entre ellos, se han detectado algunos específicos en relación con sistemas de IA, frente a los cuales se proponen instrumentos de control y defensa que se presentan como manifestación genuina del principio de responsabilidad proactiva[88].

Conforme a este principio, en efecto, al responsable del tratamiento le corresponde la comprobación del cumplimiento de todas garantías, lo que le obliga a acreditar y documentar, de forma justificada, su cumplimiento a lo largo de toda la vida del sistema. A lo que se suma la debida participación del Delegado de Protección de Datos, cumpliendo el papel que le exige el art. 38 RGPD, así como, en caso de externalización del diseño o ejecución de la solución de IA, la exigencia de cumplimiento de las obligaciones del encargado del tratamiento por parte del contratista en los términos del art. 28 RGPD y del correlativo 33 LOPD.

Este escenario no será, como advertíamos, en absoluto infrecuente. Lo determinante será en todo caso que el encargado ofrezca las garantías suficientes de que está en disposición de aplicar medidas técnicas y organizativas apropiadas que aseguren que el tratamiento que se le encarga se desarrollará en cumplimiento de las exigencias del RGPD y la LOPDGDD. Tales garantías deberán cumplir todos los parámetros que establece el art.

---

88 AEPD (2020: 42-44) toma criterio sobre todas estas cuestiones y, entre otros extremos, precisa que "los desarrolladores del componente de IA, cuando estén utilizando soluciones basadas en ML y con propósito de documentar y de cumplir con el principio de responsabilidad proactiva deben implementar otros registros, como aquellos que permiten realizar la trazabilidad sobre la procedencia de los datos de entrenamiento y validación, así como registros de los análisis que se han realizado sobre la validez de dichos datos y sus resultados".

28 RGPD –entre los que se incluyen previsiones especiales en el caso de transferencias internacionales de datos, más que comunes para los casos de alojamiento de datos "en la nube"[89]-, y así han de quedar debidamente reflejadas en el contrato que suscriba con la Administración responsable del tratamiento, con especial atención al cumplimiento de las medidas de seguridad en los términos del art. 32 RGPD, sabedores de que el encargado debe aplicar el Esquema Nacional de Seguridad, según le exige la DA 1ª LOPD. La AEPD, junto con las autoridades catalana y vasca, ha elaborado una Guía para la redacción de dichos contratos, en ejercicio de la facultad que le reconoce el apartado 9 del mismo art. 28 RGPD[90].

La posición del responsable queda reforzada toda vez que podrá someter al encargado a inspecciones o auditorías, propias o encargadas a tercero, para comprobar el cumplimiento de todas las obligaciones que le corresponden *ex* RGPD, incluso cuando haya optado por encargar a su vez parte de las actividades del tratamiento, según le permite, previa autorización del responsable, el mismo art. 28 en sus apartados 2 y 6.

Le corresponda su ejecución al responsable o al encargado, interesa ahora repasar algunas exigencias destacadas que pueden determinar, de no ser atendidas, la invalidez de los actos que se sustenten en tratamientos que nos las cumplan.

Un primer criterio determinante al respecto, y que es importante retener, es la debida existencia de base jurídica legitimadora de cada concreto tratamiento, en cada fase del ciclo de vida, en su caso. Si es una Administración la directamente responsable, típicamente encontrará cobertura en el supuesto del apartado 1.e) del art. 6, que contempla a tal efecto el necesario "cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos conferidos al responsable del tratamiento; razones de interés público o el ejercicio de poderes públicos". Bajo semejantes paraguas, la concreta base jurídica que de cobertura a un concreto tratamiento –necesariamente legal, al menos de

[89] La AEPD también ha elaborado, en 2018, una guía para estos supuestos: Orientaciones para prestadores de servicios de *cloud computing.*

[90] En esta Guía, referenciada en el anexo de Documentación, se incluye un clausulado de referencia.

forma mediata, *ex* art. 8.2 LOPDGDD- debe respetar los criterios exhaustivos del apartado 3 del mismo, así como el principio de proporcionalidad[91].

En los supuestos de datos necesarios para el entrenamiento, no será inusual que se deba requerir el consentimiento (ap. 1.a) o que pueda llegar a apelarse al criterio del interés legítimo (ap. 1.f), siendo así que si los datos se han obtenido de terceros, el responsable deberá asegurarse de la legitimidad de la base jurídica[92], sin descartar que se pueda apelar a que se trata de un uso compatible en los términos del apartado 4 del mismo art. 6.

La puesta en escena de la exigencia potencial del consentimiento del interesado da ocasión para matizar la asunción más o menos común de que los problemas de gestión de datos en relación con los sistemas de IA se mitigan por cuanto manejan típicamente datos anonimizados, y ello en la medida en que, en los términos del cdo. 26 RGPD, aquellos quedan extramuros de su ámbito de aplicación[93]. Ha de tenerse en cuenta, en todo caso, que anonimización y pseudoanonimización no son técnicas de aplicación sencilla, como demuestran las aclaraciones que ha debido efectuar al respecto la AEPD sobre la base de que esta última sí está sometida a las reglas del RGPD y de la LOPDGDD[94]. El RGPD la define, en efecto, en su art. 4.5[95], mientras que la DA 17ª de la LOPDGDD permite, bajo

---

91 Tal y como pone de manifiesto J. VALERO TORRIJOS (2023: 368-370), la norma con rango de ley es imprescindible, para establecer las condiciones jurídicas básicas del tratamiento, con referencia a los sujetos afectados, las cesiones previstas y su duración, con atención a su obligada difusión a través del registro de Actividades de Tratamiento, por más que la adaptación de la solución de IA al efectivo cumplimiento de las exigencias del RGPD puede concretarse a través de meras decisiones de los órganos competentes.

92 AEPD (2020: 20-23).

93 En este sentido A. HUERGO LORA (2022: 88).

94 La AEPD (2016) formuló unas Orientaciones y garantías en los procedimientos de anonimización de datos personales, en las que se pone de manifiesto que, por principio, ninguna técnica de anonimización puede garantizar en términos absolutos la imposibilidad de reidentificación, de modo que es imprescindible un análisis de riesgos al respecto. Más recientemente, en AEPD (2021b), formuló el documento "10 malentendidos relacionados con la anonimización", en el que advierte de los riesgos de confundir anonimización y pseudoanonimización.

95 Identificándola con el tratamiento de datos personales cuyo efecto es que ya no puedan atribuirse a un interesado sin utilizar información adicional, siempre que dicha información adicional figure por separado y esté sujeta a medidas técnicas y organizativas destinadas a garantizar que los datos personales no se atribuyan a una persona física identificada o identificable. Frente a ello, la anonimización, a falta de definición normativa, se puede entender como el proceso que permite eliminar o reducir al mínimo los riesgos de reidentificación de un individuo a partir de sus datos personales eliminando toda referencia directa o indirecta a su

ciertas condiciones, la reutilización sin consentimiento de datos pseudoanonimizados para fines de investigación en salud.

Si verdaderos datos personales están en juego, es imprescindible que el responsable del tratamiento –como proyección del principio de transparencia[96]- informe al interesado -la persona física cuyos datos personales van a ser objeto de tratamiento- del tratamiento mismo y de sus circunstancias, en función de la etapa del ciclo de vida de que se trate. En los términos del art. 13 RGPD, cuando los datos se obtengan directamente del interesado, y en los del 14, cuando se hayan obtenido indirectamente, aunque en este caso se contemplan varias excepciones que no eximen necesariamente de la adopción de las debidas cautelas. El –correlativo- derecho del interesado a ser informado ha sido concretado, respectivamente, en los apartados 1 y 2, y 3[97], del art. 11 de la LOPDGDD, distinguiendo entre información básica a facilitar directamente, y la información que debe ser puesta a disposición a través de una dirección electrónica u otro medio sencillo. Las previsiones sobre el complementario derecho de acceso que contempla el art. 15 RGPD no han encontrado especial concreción, sin embargo, en el art. 13 de la LOPDGDD[98].

De particular interés en este estudio es la precisión que se contiene, en términos idénticos, en los apartados 2.f), 2.g) y 1.h) de los tres artículos que nos ocupan, siendo así que el interesado tendrá derecho a ser informado –y, de hecho, lo será directamente en los términos del art. 11, apartados 2 y 3 LOPDGDD- y/o a acceder a información relativa a "*la existencia de* ***decisiones automatizadas, incluida la elaboración de perfiles, a que se***

identidad, pero manteniendo la veracidad de los resultados del tratamiento de los mismos, según AEPD (2020: 9).

96 Presente en los el art. 5.1.a) y en los cdos. 39, 58 y 78 en términos que van más allá de las obligaciones que pasamos a describir en el texto, sobre la base del principio de privacidad por defecto (art. 5.1.b).

97 En el caso de que los datos personales no hayan sido obtenidos directamente del interesado, a la información básica exigida para el caso contrario en el apartado 2 del art. 11, habrá que añadir información acerca de: "a) Las categorías de datos objeto de tratamiento" y "b) Las fuentes de las que procedieran los datos".

98 Al hilo de esta reflexión conviene apuntar que la AEPD (2020: 25-) pone de manifiesto que para el caso de que los datos personales se distribuyan entre una red de responsables es necesario, siguiendo el principio de responsabilidad proactiva, incluir un modelo de "gobernanza de la información" efectivo que permita la trazabilidad de la información para poder identificar al responsable y hacer posible el ejercicio de los derechos que le corresponden a los interesados. Este será el caso del propio derecho de acceso, pero igualmente el de supresión (art. 17) –especialmente relevante en relación con la fase de entrenamiento-; el de rectificación (art. 16) y el de portabilidad (art. 20).

refiere el artículo 22, apartados 1 y 4, y, al menos en tales casos, información significativa sobre la lógica aplicada, así como la importancia y las consecuencias previstas de dicho tratamiento para el interesado".

El art. 22, en su apartado 1, impone una prohibición general de adopción de decisiones basadas exclusivamente en el tratamiento automatizado de datos –sin supervisión humana alguna, según aclara la AEPD-[99], incluida la elaboración de perfiles, para el caso de que produzca efectos jurídicos en el interesado o le afecte significativamente de modo similar. Su apartado 2, sin embargo, introduce algunos supuestos tasados de excepción, entre los que resulta para nosotros especialmente pertinente, asumiendo que nos refiramos a tratamientos efectuados directamente por la Administración, el que, en el apartado b), contempla el caso de que tales decisiones estén autorizadas por el Derecho de la Unión o de los Estados miembros que se aplique al responsable del tratamiento y que establezca asimismo medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado. Dificílmente encajará la excepción del consentimiento explícito, que vuelve a aparecer en el apartado c), por cuanto la posición privilegiada de la Administración ante el ordenamiento difícilmente se concilia con la lógica del consentimiento del administrado para el tratamiento de sus datos cuando se trata de ejercer potestades públicas[100], en coherencia con lo dispuesto en el ya mencionado art. 6.1.e) RGPD.

No deja de sorprender, sin embargo, que el apartado 3 del mismo art. 22 excluya exclusivamente para el caso del apartado 2.b) la especial exigencia de que entre las medidas que se adopten para salvaguardar los derechos y libertades y los intereses legítimos del interesado se incluya "como mínimo el derecho a obtener intervención humana por parte del responsable, a expresar su punto de vista y a impugnar la decisión". Tal compromiso se impone específicamente al responsable del tratamiento, con lo que podría entenderse que para el caso del 2.b), asume que será la norma nacional o europea que posibilite este particular tratamiento la que calibre la oportunidad de introducir unas

---

99 Según precisa la AEPD (2020: 28), es imprescindible que no haya intervención humana, entendida como supervisión de la decisión por persona competente y autorizada para modificarla a través de una acción significativa y no simbólica.

100 J. VALERO TORRIJOS (2023: 364-366) pone de manifiesto, haciéndose eco del propio RGPD (cdo. 43) y de la posición de la AEPD, cuán difícil es que, siendo una autoridad pública la responsable del tratamiento, pueda hablarse de consentimiento libre.

salvaguardas que, en los dos últimos casos, son consustanciales a la actuación administrativa.

El apartado 4, al que sí remiten expresamente los arts. 13 y 14 RGPD, precisa, por su parte, que no podrá tolerarse la adopción de decisiones automáticas basadas en las categorías especiales de datos personales contempladas en el artículo 9, apartado 1, salvo que se aplique el artículo 9.2, letra a) o g) –es decir, por explícito consentimiento (poco probable, insistimos) o cuando concurran razones de interés público esencial determinadas por el Derecho europeo o interno con los condiciones que establecen una y otra letra[101]-, y siempre y cuando se hayan tomado medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado[102].

La referencia a las decisiones automatizadas nos enfrenta con un supuesto muy concreto, el más delicado, en que puede consistir el uso de sistemas de IA por parte de las Administraciones Públicas, que en el seno del RGPD se aborda, exclusivamente, en la medida en que implique, al menos por inferencia, datos personales. Sobre él volveremos extensivamente en el apartado siguiente al tratar del principio de transparencia en el marco del art. 41 LRJPAC, pero precisamente por ello interesa ahora llamar la atención sobre la precisión que contienen los arts. 13.2.f) y 14.2.g) cuando exigen que el interesado reciba "*al menos en tales casos, información significativa sobre la lógica aplicada, así como la importancia y las consecuencias previstas de dicho tratamiento para el interesado*". La AEPD relaciona esta exigencia con el concepto de "explicabilidad" del tratamiento mediante IA, poniendo énfasis en el requisito de que la información sea "significativa" para advertir que "cumplir con esta obligación ofreciendo una referencia técnica a la implementación del algoritmo puede ser opaco, confuso, e incluso conducir a la fatiga informativa". Frente a ello aboga por facilitar información que permita entender el comportamiento del tratamiento, como podría ser el caso de la que incluya el detalle de los datos empleados para la toma de decisión y la precisión de la importancia relativa otorgada a cada uno, la calidad de los datos de entrenamiento y los tipos de patrones utilizados, los perfilados realizados y sus

---

101 Hay que tener en cuenta que, tal y como recuerda la STC 76/2019, de 22 de mayo, FJ 4, el art. 9.2 LOPD señala que "deberán estar amparados en una norma con rango de ley, que podrá establecer requisitos adicionales relativos a su seguridad y confidencialidad".

102 Al respecto, vid. cdo. 71 RGPD.

implicaciones, la existencia o no de supervisión humana y, particularmente, la referencia a auditorias, especialmente sobre las posibles desviaciones de los resultados de las inferencias, así como las certificaciones realizadas al sistema de IA[103].

Estas dos piezas -auditoría y certificaciones- pueden jugar, junto con las Evaluaciones de Impacto, un importante papel como garantía y control de los sistemas de IA, con trascendencia más allá incluso de la protección de los datos personales.

Tal y como ha puesto de manifiesto la AEPD, el proceso de auditoría sobre un tratamiento de datos que incluya un componente de IA puede concentrarse sobre distintos elementos del aquel, como el proceso, la distribución del modelo y su seguridad o robustez o aspectos concretos del tratamiento o de la operación. Puede ejecutarse de forma automática o manual, interna o externa y con fines de control o transparencia. Resulta así un instrumento muy versátil para materializar el principio de responsabilidad proactiva que establece el art. 5.2 RGPD, que en último término implica, al exigir la documentación continua del tratamiento durante su ciclo de vida, la trazabilidad de los procesos y, en definitiva, su auditabilidad. La necesidad de que la auditoría se realice en las mismas condiciones que las del entorno real de explotación, obliga, en el caso de los tratamientos que integren soluciones de IA, que este extremo sea particularmente tenido en cuenta a los efectos de delimitar el contexto y el entorno a comprobar, asegurándose de comprobar que la solución de IA se está utilizando para el propósito para el que fue diseñadas[104].

El art. 42 RGPD contempla, por su parte, la posibilidad de establecer mecanismos de certificación específicos en materia de protección de datos, así como sellos de calidad, emitidos por terceros independientes. Unos y otros pueden jugar un papel promotor del cumplimiento proactivo de las obligaciones legales en la materia que permite, además,

---

103 AEPD (2020: 24).

104 AEPD (2020: 45-47) advierte que la auditoría debe realizarse con la finalidad de repasar que el tratamiento cumple los requisitos que marca la normativa en los aspectos relevantes según cual sea su objeto. En relación con los sistemas de toma de decisiones automatizadas, recomienda que se apliquen herramientas de auditoria automática en tiempo real para asegurar la coherencia y precisión de los resultados, permitiendo, en su caso, que las decisiones erróneas sean oportunamente canceladas. En 2021, publicó unas recomendaciones específicas al efecto AEPD (2021a).

conciliar las cuestiones de confidencialidad que se derivan de la protección de la propiedad industrial. Son además instrumento de transparencia.

Por lo que hace a las Evaluaciones de Impacto de Protección de Datos (EIPD) hay que tener en cuenta que aunque un concreto tratamiento no desemboque en una decisión automatizada debe ser sometido a un análisis de riesgos como proyección del mencionado principio de responsabilidad proactiva. Análisis de riesgos y gestión de riesgos son, como sabemos, dos piezas concatenadas en la regulación de riesgos que en el seno del RGPD encuentran especial manifestación en la Evaluación de Impacto contemplada en su art. 35. Esta Evaluación será preceptiva, particularmente cuando el tratamiento utilice "nuevas tecnologías", para los supuestos de tratamientos que sea probable que entrañen un "alto riesgo para los derechos y libertades de las personas físicas" teniendo en cuenta "su naturaleza, alcance, contexto o fines". Y ha de formularse con carácter previo al inicio del tratamiento bajo los parámetros de la garantía de privacidad desde el diseño y por defecto. En su elaboración se deberá recabar el asesoramiento del Delegado de Protección de Datos, en su papel clave para la debida gestión del riesgo de los tratamientos, y –esto es importante- se podrá recabar la opinión de los interesados, incluyendo los operadores humanos que interpretan o supervisan los resultados de la IA, siempre sin perjuicio de la protección de intereses públicos o comerciales –lo cual es revelador- o de la seguridad de las operaciones de tratamiento.

La AEPD, en cumplimiento del mandato del apartado 4, ha publicado una lista de los tipos de operaciones que requieren evaluación, completada con la de tratamientos que no la requieren, según posibilita el apartado 5. Conforme al planteamiento de la primera, no exhaustiva[105], es más que probable que la evaluación sea necesaria en la mayoría de los casos de aplicación de soluciones de IA[106].

---

105 Esta lista se basa en los criterios establecidas por el Grupo de Trabajo del Artículo 29 en la Guía WP248 "Directrices sobre la evaluación de impacto relativa a la protección de datos (EIPD) y para determinar si el tratamiento «entraña probablemente un alto riesgo» a efectos del RGPD", de la que debe considerarse complementaria.

106 Cuando se den al menos dos de sus criterios la evaluación es necesaria, siendo así que entre ellos se refiere expresamente a los "tratamientos que impliquen la toma de decisiones automatizadas o que contribuyan en gran medida a la toma de tales decisiones, incluyendo cualquier tipo de decisión que impida a un interesado el ejercicio de un derecho o el acceso a un bien o un servicio o formar parte de un contrato" (ap. 2) y a los "que impliquen el uso de datos a gran escala", a cuyo fin se tomarán en cuenta los criterios establecidos en la guía WP243

Resulta, con todo, que en el listado de supuestos que no requieren Evaluación, igualmente orientativo[107], se incluye una previsión especialmente pertinente en relación con la actividad de las Administraciones Públicas, por cuanto excluye de tal exigencia los "tratamientos que sean necesarios para el cumplimiento de una obligación legal, cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos conferidos al responsable", si bien con la condición –un tanto paradójica- de que "en el mismo mandato legal no se obligue a realizar una EIPD, y siempre y cuando ya se haya realizado una EIPD completa". La paradoja cobra sentido a la luz de lo previsto en el apartado 10 del mismo art. 35, que contempla por sí mismo un supuesto de exoneración de la evaluación previa al tratamiento (a salvo que los Estados miembros la consideren necesaria), en los casos de tratamiento de los previstos en los apartados c) y e) del art. 6.1 –precisamente los enunciados en el listado en el supuesto que nos ocupa-, cuando durante el proceso de adopción de la base jurídica en el Derecho de la Unión o en el Derecho del Estado miembro que se aplique al responsable del tratamiento, se hubiera sometido a una evaluación de impacto relativa a la protección de datos como parte de una evaluación de impacto general. Y ello siempre y cuando tal base jurídica hubiera regulado la operación específica de tratamiento o conjunto de operaciones en cuestión, como es obvio, siendo este momento clave, en puridad, para valorar la oportunidad misma de introducir un sistema de IA.

La Evaluación de Impacto será, en definitiva -en sentido propio, o por elevación-, instrumento usual en la órbita de las actuaciones de las Administraciones Públicas que apliquen IA[108]. Incluso más allá de la obligación *ex* RGPD, tal y como recomienda la AEPD en su documento "Gestión de riesgos y evaluación de impacto en tratamiento de

"Directrices sobre los delegados de protección de datos (DPD)" del Grupo de Trabajo del Artículo 29. También se enuncian los que supongan la elaboración de perfiles; los que determinen la observación, monitorización, supervisión, geolocalización o control de personas de forma sistemática y exhaustiva; los que impliquen el uso de categorías especiales de datos a los que se refiere el art. 9.1, en particular biométricos; los que impliquen la asociación, combinación o enlace de registros de bases de datos de dos o más tratamientos con finalidades diferentes o por responsables distintos; los que afecten a sujetos vulnerables o en riesgo de exclusión social o los que supongan la utilización de nuevas tecnologías o un uso innovador de tecnologías consolidadas.

[107] Lista orientativa de tipos de tratamientos que no requieren una evaluación de Impacto relativa a la protección de datos según el artículo 35.5 RGPD, sin fecha.

[108] De hecho, la AEPD pone a disposición una plantilla para la elaboración de la Evaluación particularmente por parte de entidades del sector público.

datos personales", de junio de 2021[109], al defender su conveniencia extensiva en su condición de herramienta clave dentro de la visión global de la gestión de riesgos –altos o no- que le compete asumir al responsable del tratamiento bajo criterios de plena responsabilidad, valga el juego de palabras, y no como un mero ejercicio pro forma[110]. Dicho lo cual, es indudable que cobra una dimensión más intensa en los casos de alto riesgo, para los que incluso puede requerirse la consulta a la AEPD previa al tratamiento en los términos del art. 36, sobre la base de las conclusiones que arroje la propia Evaluación respecto determinados riesgos.

Entre ellos, particularmente, los intrínsecos a los tratamientos que se basen en soluciones de IA para la toma de decisiones, respecto de los que la AEPD ya ha recomendado que la EIPD lleve a cabo un análisis comparativo entre el rendimiento obtenido por un operador humano cualificado frente a los resultados arrojados por modelos automáticos, teniendo en cuenta las condiciones reales de entrada y el contexto en el que se despliega el tratamiento[111]. La prudencia en la implantación de soluciones de IA vuelve a ser, pues, la mejor receta, máxime si somos conscientes de las enormes exigencias prácticas que implica la aplicación de los mecanismos preventivos que venimos describiendo, lo que lastra inevitablemente la verosimilitud de su eficacia real[112].

B.- Especial atención a los sesgos y a sus circunstancias en el seno de la Ley 15/2022

---

109 AEPD (2021c). Tal y como argumenta J. VALERO TORRIJOS (2023: 383), el carácter meramente orientativo de este documento impide derivar de su incumplimiento la no conformidad a Derecho de una evaluación que no cumpliera sus criterios.

110 Téngase en cuenta que, en los términos del apartado 7 del mismo art. 35, que la evaluación deberá incluir como mínimo a) una descripción sistemática de las operaciones de tratamiento previstas y de los fines del tratamiento, inclusive, cuando proceda, el interés legítimo perseguido por el responsable del tratamiento; b) una evaluación de la necesidad y la proporcionalidad de las operaciones de tratamiento con respecto a su finalidad; c) una evaluación de los riesgos para los derechos y libertades de los interesados a que se refiere el apartado 1, y d) las medidas previstas para afrontar los riesgos, incluidas garantías, medidas de seguridad (lo que supondrá la aplicación del Esquema Nacional de Seguridad, conforme el art. 32 RGPD y DA 1ª LOPD, según razonamos supra) y mecanismos que garanticen la protección de datos personales, y a demostrar la conformidad con el RGPD, teniendo en cuenta los derechos e intereses legítimos de los interesados y de otras personas afectadas. De forma añadida, hay que tener en cuenta que en los términos del apartado 11, el responsable examinará, en caso necesario, si el tratamiento es conforme con la evaluación de impacto relativa a la protección de datos, al menos cuando exista un cambio del riesgo que representen las operaciones de tratamiento. En AEPD (2020: 32-33) se ofrecen detalles sobre el contenido deseado de la Evaluación.

111 AEPD (2020: 44).

112 Algún apunte en este sentido en A. SORIANO ARNAIZ (2021).

El RIA se muestra también muy atento, como advertíamos, a las amenazas que para el principio de igualdad pueden implicar los sistemas de IA cuando, basándose en el manejo masivo de datos, si estos no están debidamente depurados, generan sesgos, deliberados o no. Los sesgos en los modelos de inferencia están íntimamente ligados con la calidad del dato, que es reflejo del principio de exactitud que enuncia el art. 5.1.d RGPD[113]. Relevante es la distinción entre datos duros y blandos, donde los primeros son objetivos, generalmente cuantificables, mientras que los segundos integran un componente subjetivo o de incertidumbre, siendo estos últimos, evidentemente, los más propensos a facilitar sesgos.

Se consideran sesgos cualesquiera anomalías en la información de salida del sistema de IA debidas a prejuicios o suposiciones erróneas realizadas durante el proceso de desarrollo del sistema o determinados por los datos de entrenamiento, con la consecuencia de que sus resultados no son generalizables ampliamente, y pueden producir, directa o indirectamente, efectos discriminatorios[114]. Son tres factores los que pueden influir en la generación de sesgos: la propia programación o diseño del sistema de IA, los datos de entrenamiento y validación y la evolución del modelo, siendo especialmente relevante la segunda[115]. Resulta por todo ello imprescindible un proceso de corrección permanente de los sesgos que se detecten[116].

---

113 AEPD (2020: 35-38) ofrece varias precisiones respecto del contenido y alcance de este principio en relación con la generación de sesgos.

114 Esta definición glosa la que ofrecen M. ESTEVEZ ALMENZAR et al (2022: 16-19). Para esta referencia y otras varias, L. COTINO HUESO (2023: 260-261), quien igualmente se hace eco de la clasificación que ofrecen las "Guías Éticas para una IA fiable", ya citadas, que distinguen entre sesgos sistémicos, sesgos estadísticos y computacionales y sesgos humanos. La AEPD (2020: 7), por su parte, ofrece una sencilla definición: "desviación inadecuada en el proceso de inferencia".

115 Téngase en cuenta que, particularmente en los supuestos de *Machine Learning*, los datos obtenidos son sometidos a procesos de revisión, limpieza y transformación previamente a ser sometidos al sistema de Minería de Datos propiamente dicho, en el que se aplica el algoritmo de aprendizaje. El riesgo de sesgos requiere una depuración en esta fase de obtención y puesta a disposición de los datos, si bien hay técnicas para evitar el sesgo en los conjuntos de datos. Con todo, tal y como ponen de manifiesto F. GONZÁLEZ CABANES y N. DIAZ DIAZ (2023: 69-70), la tendencia se dirige a la puesta a disposición de datos de mayor calidad para facilitar la extracción de información de los mismos, planteamiento que solo tiene sentido en la fase de entrenamiento.

116 La amenaza de los sesgos es cuestión recurrente en la literatura sobre IA. J. PONCE (2018) describe supuestos, dificultades y técnicas para su detección y corrección. La clave estará,

En el art. 10.2.f) y, correlativamente, en el apartado 5 del mismo artículo, el RIA permite a los proveedores, como medida excepcional para asegurar la vigilancia, detección y corrección de los sesgos asociados a estos sistemas –y solo en la medida en que sea estrictamente necesario-, que traten las categorías de datos personales cuyo tratamiento queda, como regla, prohibido, por el art. 9.1 RGPD, es decir, aquellos "que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas, o la afiliación sindical, y el tratamiento de datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona física[117], datos relativos a la salud o datos relativos a la vida sexual o las orientación sexuales de una persona física"[118]. Es condición para ello, como no podía ser de otro modo, que se ofrezcan las salvaguardias adecuadas para los derechos y las libertades fundamentales de las personas físicas, lo que incluye establecer limitaciones técnicas a la reutilización y la utilización de las medidas de seguridad y protección de la privacidad más recientes, tales como la seudonimización o el cifrado, cuando la anonimización pueda afectar significativamente al objetivo perseguido[119].

---

en todo caso, tanto o más que en la selección original de los datos a manejar, en una depuración continúa de los resultados no deseados que se detecten.

[117] Estos datos, definidos en el art. 4.14 del mismo RGPD, como "datos personales obtenidos a partir de un tratamiento técnico específico, relativo a las características físicas, fisiológicas o conductuales de una persona física que permitan o confirmen la identificación única de dicha persona, como imágenes faciales o datos dactiloscópicos", serán, como hemos visto, resultado de algunos sistemas de IA, directamente prohibidos, o declarados de alto riesgo según la Ley europea de IA, de modo que en este último caso deberá tenerse presente el planteamiento del art. 9 RGPDA, muy restrictivo de su posible tratamiento. Con todo, la obtención de datos biométricos y su manejo para funciones públicas, como pueda ser el control horario de los empleados públicos, puede resultar perfectamente acomodado a la normativa de protección de datos, como se constató de hecho en la STS de 2 de julio de 2007 (Roj: STS 5200/2007).

[118] El precepto también cita a tales efectos el art. 10 de la ya citada Directiva (UE) 2016/680, y el art. 10.1 del Reglamento (UE) 2018/1725, del Parlamento Europeo y del Consejo, de 23 de octubre de 2018, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por las instituciones, órganos y organismos de la Unión, y a la libre circulación de esos datos, y por el que se derogan el Reglamento (CE) nº 45/2001 y la Decisión nº 1247/2002/CE. Siendo así que ambos preceptos contemplan la posibilidad de excepcionar, aun bajo restricciones y garantías adicionales, las prohibiciones que también establecen para el tratamiento de tales datos dentro de su ámbito de aplicación. Hay que entender que tales restricciones y garantías específicas serán aplicables cuando los correspondientes sistemas de IA de alto riesgo se vayan a aplicar en el ámbito de una u otra normas.

[119] Estas previsiones de la Ley de IA vienen a materializar la opción de excepción a la regla de prohibición de tratamiento que contiene el apartado 2.b) del mismo art. 9.

El art. 15.3, por su parte, hace una llamada de atención específica para asegurar la solidez de los sistemas que continúen aprendiendo tras su puesta en explotación en punto a la posible aparición de sesgos por efecto del bucle de retroalimentación, es decir, los debidos al manejo de la información de salida como información de entrada sucesiva para el aprendizaje. Para tales casos, sin mayor precisión, se requiere la introducción de medidas de mitigación oportunas.

La propia Ley 15/2022, en su art. 25, obliga a aplicar métodos o instrumentos suficientes para la detección, la adopción de medidas preventivas, y la articulación de medidas adecuadas para el cese de las situaciones discriminatorias como las que puedan implicar los sesgos derivados de los sistemas de IA, siendo así que el incumplimiento de tales obligaciones dará lugar a responsabilidades administrativas, así como, en su caso, penales y civiles por los daños y perjuicios que puedan derivarse, que podrán incluir –especifica el apartado 2- "tanto la restitución como la indemnización, hasta lograr la reparación plena y efectiva para las víctimas". Deberán, además, tomarse las medidas para que no vuelvan a repetirse "incidentes discriminatorios", "especialmente en los casos en los que el agente discriminador sea una administración pública", según precisa el apartado 3 del mismo art. 25. El art. 26, por su parte, en sintonía con el art. 47.1.a) LRJPAC, hace expreso el carácter nulo de pleno Derecho de las disposiciones, actos o cláusulas de los negocios jurídicos que constituyan o causen discriminación por razón de alguno de los motivos previstos en el art. 2.1 de la misma Ley, que reconoce a toda persona -con independencia de su nacionalidad, mayor o menor edad, condición o no de residente legal- el derecho a la igualdad de trato y no discriminación por razones bien amplias: nacimiento, origen racial o étnico, sexo, religión, convicción u opinión, edad, discapacidad, orientación o identidad sexual, expresión de género, enfermedad o condición de salud, estado serológico y/o predisposición genética a sufrir patologías y trastornos, lengua, situación socioeconómica, o cualquier otra condición o circunstancia personal o social.

La depuración de los sesgos derivados de sistemas de IA utilizados por Administraciones públicas podrá encontrar cauce a través del mecanismo del art. 31 de la misma Ley, que impone a las autoridades públicas que, con ocasión del ejercicio de sus competencias, tengan conocimiento de un supuesto de discriminación de los previstos en esta ley,

incoen, si son competentes, el correspondiente procedimiento administrativo, “en el que se podrán acordar las medidas necesarias para investigar las circunstancias del caso y adoptar las medidas oportunas y proporcionadas para su eliminación”. De no ser competentes, deberán comunicar los hechos de forma inmediata a la Administración competente. Crucial será al respecto el reconocimiento de la condición de interesado en los correspondientes procedimientos que el apartado 2 del mismo artículo atribuye a los sindicatos, las asociaciones profesionales de trabajadores autónomos, las organizaciones de personas consumidoras y usuarias y a las asociaciones y organizaciones legalmente constituidas que tengan entre sus fines la defensa y promoción de los derechos humanos y cumplan los requisitos fijados en el artículo 29.2, siempre que cuenten con la autorización de las personas afectadas o sin falta de ella cuando las personas afectadas sean una pluralidad indeterminada o de difícil determinación, sin perjuicio de que quienes se consideren afectados puedan también participar por sí mismos en el procedimiento.

Algún papel podría jugar también la recién creada Autoridad Independiente para la Igualdad de Trato y la No Discriminación, a la que el art. 40.c) le atribuye competencia para iniciar, de oficio o instancia de terceros, investigaciones sobre la existencia de posibles situaciones de discriminación que revistan una especial gravedad o relevancia por razón de las causas previstas en el citado art. 2.1, siempre y cuando no consistan en una presunta infracción penal. La Autoridad podrá, también, ejercitar acciones judiciales en defensa de los derechos derivados de la igualdad de trato y la no discriminación.

Llegados a este punto, y para concluir el repaso de las medidas de minimización de riesgos que es conveniente contemplar para la implantación de soluciones de IA por parte de las Administraciones públicas, es preciso apuntar alguna breve reflexión sobre una cuestión clave que se ha suscitado ya en más de un momento: la de la necesidad o no, y en qué grado, de supervisión humana del funcionamiento de los sistemas.

4.- Las garantías de vigilancia humana y su proyección sobre la inteligibilidad del sistema en aras de su transparencia. Remisión.

Aunque a día de hoy, solo los sistemas de IA más avanzados son capaces de operar sin supervisión humana, es, en último término, una decisión de diseño del sistema el dotarle

o no de autonomía[120]. De ahí que resulten pertinentes las previsiones que el art. 14 del RIA contiene para asegurar la vigilancia humana de los sistemas de IA de alto riesgo, prescribiendo desde su apartado 1 que estos sistemas se diseñarán y desarrollarán de modo que puedan ser vigilados de manera efectiva por personas físicas durante el período que estén en uso, lo que incluye dotarlos de una herramienta de interfaz humano-máquina adecuada, entre otras cosas, aun cuando se limite el objetivo de tal supervisión humana a la prevención o reducción al mínimo de los riesgos para la salud, la seguridad o los derechos fundamentales que pueden surgir con el uso del sistema, particularmente cuando dichos riesgos persisten a pesar de aplicar otros requisitos establecidos en el mismo capítulo (ap. 2).

En los términos del apartado 3, las concretas medidas de vigilancia humana quedan a criterio del proveedor, pudiendo quedar integradas en el propio sistema, si es que es viable técnicamente, y/o quedar derivadas a su aplicación por el usuario. Lo determinante será, con todo, el alcance de esa vigilancia humana, que conforme a las pormenorizaciones del apartado 4 será muy amplia, asegurando una serie de extremos que resultarán, particularmente en algún caso, especialmente relevantes cuando se trata de ejercitar potestades públicas a través o con apoyo del sistema de IA.

La cautela primordial pasa por introducir medidas que permitan a los humanos supervisores "ser conscientes de la posible tendencia a confiar automáticamente o en exceso en la información de salida generada por un sistema de IA de alto riesgo" –el llamado «sesgo de automatización»-, en particular con aquellos sistemas que se utilizan para aportar información o recomendaciones con el fin de que personas físicas adopten una decisión (ap. 4.b). Más aun, debemos precisar, cuando el sistema mismo se pudiera plantear como directamente decisorio.

Se trata, en definitiva, de evitar lo que la AEPD identifica con un sesgo humano consistente en aceptar sin espíritu crítico los resultados de una IA como ciertos e inamovibles, atribuyéndoles un principio de autoridad consustancial a la tecnología. Incurriendo en el dataísmo, en fin, siguiendo la expresión de Harari. Utilizando las propias palabras de la Agencia, "como buena práctica, y más allá de exigencias derivadas de la protección de datos, la supervisión humana ha de ser una opción en tratamientos

[120] Así lo puntualizan E. GÓMEZ et al (2023:731-732).

basados en el IA y en general de decisiones automatizadas. Hay que evitar el diseño de sistemas con orientación "palanca de hombre muerto" y dar siempre la opción de que un operador humano pueda ignorar el algoritmo en un momento dado, procedimentalizando aquellas situaciones en las que debe optarse por este modo de actuar. Para ello es recomendable documentar las incidencias o los cuestionamientos de las decisiones automáticas recibidas de los interesados, de modo que de su análisis sea posible detectar situaciones en las que es necesaria la intervención humana porque el tratamiento puede no estar funcionando de la manera esperada"[121].

Este escenario de funcionamiento erróneo es el que justifica el supuesto del apartado 4.a), conforme al cual se exige que los sistemas de alto riesgo permitan a las personas a quienes se encomiende la vigilancia humana "entender por completo las capacidades y limitaciones del sistema de IA de alto riesgo y controlar debidamente su funcionamiento, de modo que puedan detectar indicios de anomalías, problemas de funcionamiento y comportamientos inesperados y ponerles solución lo antes posible". Se garantiza, pues, la inteligibilidad del funcionamiento del sistema con el fin de permitir, en último término, intervenir en el mismo, hasta el punto de interrumpir el sistema accionando un botón específicamente destinado a tal fin o mediante un procedimiento similar (ap. e).

Se trata, pues, en esta primera aproximación, de permitir detectar fallos del sistema con el fin de reaccionar frente a ellos. Pero los riesgos no acaban ahí, evidentemente. Aun cuando el sistema actúe correctamente, la supervisión humana es necesaria para reaccionar frente a resultados anómalos, para lo cual es crucial introducir medidas que aseguren a los supervisores humanos la inteligibilidad de la información de salida del sistema, permitiéndoles interpretarla correctamente en función de sus características y "las herramientas y los métodos de interpretación disponibles" (ap. c).

De nuevo en palabras de la AEPD, hay que evitar que el supervisor humano actúe como una "mera correa de transmisión", por lo que hay que prevenir "errores de interpretabilidad. Los valores inferidos han de presentarse de forma que reflejen la realidad de la inferencia y sus límites a los operadores humanos o en las fases posteriores del tratamiento. Durante la explotación es necesario ofrecer información en tiempo real de los valores de exactitud y/o calidad de la información inferida en cada momento, de

[121] AEPD (2020: 8 y 28-29).

forma que cuando no alcance un umbral mínimo se ha de indicar explícitamente que es nula o no tiene ningún valor".

En este planteamiento encaja la precisión del apartado 4.d), que obliga a que se asegure a los supervisores humanos la posibilidad de "decidir, en cualquier situación concreta, no utilizar el sistema o desestimar, invalidar o revertir la información de salida que este genere". La regla es, pues, la supervisión reactiva, que permita descartar los resultados anómalos frente a un automatismo ciego. En este escenario es crítico el sesgo humano, no así en el modelo reforzado para los sistemas de identificación biométrica que introduce el apartado 5, conforme al cual, y bajo una lógica preventiva, se exige que se garantice que el usuario no actúe ni tome ninguna decisión sobre la base de la identificación generada por el sistema salvo verificación y confirmación de un mínimo de dos personas físicas (ap. 5).

Sobre todas estas apreciaciones habrá ocasión de volver en el apartado siguiente, en el que pasamos a abordar las garantías de transparencia que el mismo art. 23 de la Ley 15/2022 igualmente reclama, como nos consta, para los sistemas de IA.

IV.- Garantías de transparencia a favor de los ciudadanos cuando los sistemas de IA son utilizados por Administraciones Públicas: variantes en el cuándo y el cómo

1.- Algunas precisiones sobre el carácter gradual de las exigencias de transparencia en función del ámbito de aplicación de las soluciones de IA por Administraciones públicas

Retomando los términos del art. 15 de la Ley 22/2023, las Administraciones públicas favorecerán la puesta en marcha de mecanismos para que los algoritmos involucrados en la toma de decisiones que se utilicen en las administraciones públicas tengan en cuenta criterios de transparencia, además de minimización de sesgos –de riesgos hemos analizado, por extensión- y de rendición de cuentas, "siempre que sea factible técnicamente". Pone así en escena el precepto que hemos tomado como referencia un extremo de capital importancia en la órbita de las garantías exigibles para asumir la aplicación de sistemas de IA por parte de las Administraciones Públicas. De su importancia da cuenta, de hecho, que el apartado 2 insista en que las Administraciones

públicas priorizarán la transparencia en el diseño y la implementación y la capacidad de interpretación de las "decisiones adoptadas por algoritmos", dice expresamente.

El precepto se refiere acotadamente "a los algoritmos involucrados en la toma de decisiones" por parte de las Administraciones Públicas. Aunque las exigencias de transparencia deben trascender a cualquier sistema de IA que utilicen Administraciones Públicas, y abarcar todo lo relativo al sistema y no solo a los algoritmos utilizados, tales exigencias adquieren, sin duda, una especial densidad cuando los sistemas se apliquen en relación con la toma de decisiones: para sustentarlas o, con más razón, para adoptarlas.

Si los sistemas de IA se utilizan en actividades auxiliares, sin involucrar el ejercicio de potestades administrativas con proyección directa sobre los ciudadanos, es obvio que el impacto que pueda tener el uso de sistemas de IA se mitiga y, consecuentemente, las razones para exigir criterios de transparencia en su uso. Esté será típicamente el caso del manejo de chat-bots y tecnologías similares en los servicios de información por parte de las distintas Administraciones, aun cuando en tales supuestos se impondrá verosímilmente la exigencia del art. 52.1 RIA, que obliga a los operadores de sistemas de IA destinados a interactuar con personas físicas a garantizar en su diseño y desarrollo que estas estén informadas de que están actuando con un sistema de IA, salvo que este extremo resulte evidente a partir del contexto y circunstancias de utilización.

Tampoco implica consecuencias directas sobre la esfera jurídica de los ciudadanos el uso de sistemas de IA en la órbita de acciones inspectoras o de investigación, del cual hemos descrito ejemplos. Solo cuando se abra propiamente el correspondiente expediente, típicamente sancionador, se activarán las exigencias plenas del procedimiento administrativo. Es más. El debido sigilo podría hacer contraproducente informar de la aplicación de determinadas soluciones a tal fin. Es de advertir, con todo, que la introducción de estos sistemas puede permitir introducir criterios de homogeneidad y racionalidad en las decisiones de identificación de las situaciones a perseguir conforme a un planteamiento planificador que, por definición, es selectivo: de este se podrían sumar

argumentos para contra-argumentar, caso a caso, la consolidada doctrina constitucional que afirma que no hay igualdad en la ilegalidad[122].

La transparencia que pueda llegar a exigirse a estas actuaciones no sería en ningún caso comparable con la que requiere la aplicación de sistemas de IA en el ejercicio de potestades administrativas, particularmente decisorias. Y aun en ese espacio, con distinto alcance. Si partimos del concepto clásico de acto administrativo, la adopción de actos que supongan declaraciones de deseo puede apoyarse en sistemas de IA, más que ser materializada por los mismos, pero en todo caso nos encontraríamos en una esfera parecida, en cuanto al efecto del acto, que la que supone la situación de ejercicio de facultades de inspección o investigación. Si el acto consiste en una declaración de conocimiento, el sistema de IA podría aplicarse de forma directa, colocándonos con naturalidad en la esfera de la regulación de las actuaciones administrativas automatizadas contenida en el art. 41 LRJSP, sobre cuya aplicabilidad en relación con la IA se hace, por ese mero dato, ineludible pronunciarse. Planteamiento este que no hace sino aumentar su intensidad cuando se proyecta sobre las declaraciones de juicio –máxime si desemboca en un informe preceptivo- y, más aun, sobre las declaraciones de voluntad.

Es, en efecto, en relación con los actos decisorios y, en menor medida, de informe, con los que surgen las mayores dudas sobre la posibilidad misma de aplicar para su adopción sistemas de IA, con plenitud o como mero apoyo para la adopción por humanos de los acuerdos propiamente dichos. Y es en este escenario en el que, tal y como hemos tenido ocasión de anticipar y ahora debemos sustanciar, hay que calibrar la verosimilitud de las opciones reales de la IA en las actuaciones públicas. Tanto en el estadio actual de la tecnología, como en su presumible o potencial evolución. Y ello por cuanto en este ámbito –y así lo hemos anticipado- la verosimilitud de la aplicabilidad de los sistemas de IA en la esfera de la Administración Pública depende de la verosimilitud de las garantías de que estén o puedan llegar a estar revestidas -en el ordenamiento vigente, en el proyectado y en el posible- pues estas operan en la esfera interna, incluso nuclear, de la ordenación de la utilización de soluciones de IA en la actividad administrativa.

---

[122] A. HUERGO LORA (2022: 85, en nota 9) cita la STS de 19 de febrero de 2020 (Rec. 240/2018) que, en sentido inverso, descarta que exista un derecho subjetivo a no ser investigado al margen de los planes de inspección. La doctrina, insatisfactoria, de la no igualdad en la ilegalidad se formula, entre otras muchas, en la STS de 11 de marzo de 2021 (rec. 347/2019)

Cuando de la garantía de transparencia se trata, hay que distinguir tres esferas, siempre sobre la base de que lo que se trata en último término es de garantizar la transparencia de los sistemas de IA respecto a los ciudadanos afectados por el cuándo y el cómo los utilicen las Administraciones Públicas: En primer lugar, la garantía del conocimiento mismo de que se está aplicando un sistema de IA; en segundo lugar, la garantía de comprensión de cómo funciona el concreto sistema de IA por parte de la Administración que lo utiliza y, en tercer lugar, la garantía de inteligibilidad de los resultados del sistema para sus destinatarios, particularmente cuando su producto sea un acto administrativo, máxime si es decisorio, incluyendo la garantía del derecho de acceso a sus elementos técnicos, como el código fuente o el algoritmo mismo[123]. Distinguir estos tres planos es importante porque es frecuente confundir su sentido y alcance en la órbita de lo que se viene en llamar "explicabilidad" de los sistemas.

2.- Alcance de la obligación de publicidad de la utilización de sistemas de IA por parte de las Administraciones Públicas: el art. 41 LRJPAC como referente

Una primera manifestación del principio de transparencia en la utilización de soluciones de IA por las Administraciones Públicas -la más elemental- es la que consiste en dar publicidad al uso de tal solución y a sus circunstancias.

A estos efectos, y en coherencia con las reflexiones con las que iniciábamos este apartado acerca del alcance con el que este principio se ha de manifestar en función del tipo de actuaciones de que se trate, es oportuno traer a colación la regulación que el art. 41 LRJSP contiene acerca de lo que califica, y define, como "actuaciones administrativas automatizadas".

Conforme a la definición legal, "se entiende por actuación administrativa automatizada, cualquier acto o actuación realizada íntegramente a través de medios electrónicos por una Administración Pública en el marco de un procedimiento administrativo y en la que no haya intervenido de forma directa un empleado público". Precisión esta última que parece descartar los supuestos en los que exista cualquier tipo de supervisión humana, si bien sus

123 G. VESTRI (2021), por su parte, distingue los planos del conocimiento mismo de la existencia del sistema, de su funcionamiento y de la explicabilidad de sus resultados.

antecedentes y referentes permiten defender que el precepto contempla supuestos en los que la supervisión no se produce necesariamente para validar cada caso[124].

Resulta, pues, verosímil, asumir que el artículo se aplica directamente a los supuestos en los que se utilicen sistemas de IA para producir actos, sean decisorios, de trámite o de comunicación, dentro de un procedimiento administrativo[125]. Incluso aunque no contemplen aquellos supervisión humana siempre y en todo caso a caso, durante su aplicación o para validar su resultado. No es este, como sabemos, un requisito consustancial a todos los sistemas de IA, ni en particular lo es necesariamente para todos los sistemas de IA que utilicen las Administraciones Públicas –salvo que sean de alto riesgo conforme al RIA-, aunque –como venimos defendiendo, e insistiremos oportunamente- resulte recomendable en condiciones pautadas[126].

El precepto introduce, en su apartado 2, una serie de cautelas al exigir para los casos de actuaciones administrativas automatizadas que se establezca previamente "el órgano u órganos competentes, según los casos, para la definición de las especificaciones, programación, mantenimiento, supervisión y control de calidad y, en su caso, auditoría del sistema de información y de su código fuente. Asimismo, se indicará el órgano que

---

124 Así lo hacía expreso, de hecho, la definición que se presenta como precedente inmediato, la contenida en el Anexo a la Ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los Servicios Públicos, que identificaba como actuación administrativa automatizada la "actuación administrativa producida por un sistema de información adecuadamente programado sin necesidad de intervención de una persona física en cada caso singular (sic)". Para afirmar a continuación que "incluye la producción de actos de trámite o resolutorios de procedimientos, así como de meros actos de comunicación". En términos muy parecidos se pronuncia el Anexo a 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 al definir la "actuación judicial automatizada", mientras que su art. 42 introduce exigencias casi idénticas al artículo que nos ocupa.

125 El art. 44 de la Ley 26/2010, de 3 de agosto, de régimen jurídico y de procedimiento de las administraciones públicas de Cataluña es elocuente cuando precisa en su apartado 1 que "las administraciones públicas catalanas pueden realizar actuaciones automatizadas para constatar la concurrencia de los requisitos que establece el ordenamiento jurídico, declarar las consecuencias previstas, adoptar las resoluciones y comunicar o certificar los datos, actos, resoluciones o acuerdos que consten en sus sistemas de información, mediante la utilización del sistema de firma electrónica que determinen".

126 GAMERO CASADO (2023: 403-404) asume que todo desarrollo de IA se sustenta de uno u otro modo en la actuación automatizada. La clave está, sin embargo, en nuestra opinión, a la inversa, en la aplicabilidad o no de soluciones de IA para ejecutar actuaciones automatizadas. El autor ofrece, con todo, una cita virtualmente exhaustiva, a la que nos remitimos, de la ya amplia literatura jurídica al respecto, sin perjuicio de destacar el estudio pionero de I. MARTÍN DELGADO (2009), que por su fecha analizó el concepto tal y como fue perfilado en la Ley 11/2007, en términos que transcribiremos en la nota siguiente.

debe ser considerado responsable a efectos de impugnación". Aun con sus imperfecciones, el precepto obliga, en definitiva, a pautar los –no pocos- extremos técnicos y organizativos conforme a los cuales se va a desenvolver la actuación administrativa automatizada, si bien resulta especialmente equívoco respecto de la determinación de cómo se decide automatizar una actuación y de a quién se atribuye orgánicamente la misma.

Para el ámbito estatal, el art. 13.2 del Real Decreto 203/2021, de 30 de marzo, por el que se aprueba el Reglamento de actuación y funcionamiento del sector público por medios electrónicos (RAFESP), establece que "la determinación de una actuación administrativa como automatizada se autorizará por resolución del titular del órgano administrativo competente", al que, en buena lógica, hay que atribuir la actuación automatizada a todos los efectos, lo que se corresponde con las exigencias que el art. 42 LRJSP establece en punto a la firma electrónica[127]. El carácter meramente instrumental del sistema de IA facilita que se entienda así[128], y así se corrobora a partir de referentes claros como el art. 96.3 de la Ley 58/2003, de 17 de diciembre, General Tributaria[129] y en normas administrativas autonómicas como la Ley catalana 26/2010[130]. Nada ha de impedir, sin embargo, como ocurre –por ejemplo- en el Ayuntamiento de Madrid, que un órgano

---

127 Aunque el precepto concibe que se firme con el sello electrónico de la Administración pública correspondiente, amén del "órgano, organismo público o entidad de derecho público, se ha defendido con buen criterio que debe ser el sello del concreto órgano al que se atribuya la actuación automatizada. En este sentido, E. GAMERO CASADO (2023: 410).

128 Incluso en el caso de que el sistema de IA llegara a tener carácter decisorio, porque en tal caso el órgano administrativo decide hacer suyo el resultado del sistema, en su caso con vigilancia humana, tal y como razonaremos. Las conclusiones de I. MARTÍN DELGADO (2009, 366) en relación con las actuaciones administrativas automatizadas son extrapolables sin dificultad a la aplicación de sistemas de IA.

129 Que versa como sigue: "3. Los procedimientos y actuaciones en los que se utilicen técnicas y medios electrónicos, informáticos y telemáticos garantizarán la identificación de la Administración tributaria actuante y el ejercicio de su competencia. Además, cuando la Administración tributaria actúe de forma automatizada se garantizará la identificación de los órganos competentes para la programación y supervisión del sistema de información y de los órganos competentes para resolver los recursos que puedan interponerse".

130 Se trata del art. 44.3 de la Ley catalana 26/2010, de 3 de agosto, de régimen jurídico y de procedimiento de las administraciones públicas de Cataluña, que prescribe que "la actuación administrativa automatizada no afecta a la titularidad de la competencia de los órganos administrativos ni a las competencias atribuidas para la resolución de los recursos administrativos".

centralice la decisión de automatizar una actuación, en su caso a instancia del órgano competente, sin que se deje de atribuir a este la actuación automatizada[131].

¿Hay que entender que es el mismo el que determina los parámetros que el precepto exige que se precisen? ¿Con qué forma?

En este contexto no se puede obviar la polémica suscitada en torno al carácter reglamentario del algoritmo[132]. Polémica que hay que, sin embargo, trascender, al menos formulada en sus propios términos, por cuanto, como describimos en su momento, el algoritmo no es en sí mismo el elemento único, ni siquiera necesariamente principal de los sistemas de IA, a lo que se suma que, en los más de los casos, no es necesariamente diseñado *ad hoc* para su utilización por la Administración, menos aun involucrándose directamente en su diseño. La mayoría de los algoritmos que se utilizan están a disposición, incluso en código abierto, en bibliotecas, como ocurre en general en la actualidad con todos los elementos de programación. Salvo para aplicaciones de alta seguridad, no se diseña desde 0 ningún software, sea o no de IA.

Cosa distinta es que sea oportuno y necesario que la Administración que decida utilizar un sistema de IA, particularmente si lo va a utilizar en actuaciones procedimentalizadas como las que acota el art. 41 LRJSP, esté obligada a fijar previamente las "especificaciones", "programación", "mantenimiento", "supervisión y control de calidad" y, en su caso, "auditoría del sistema de información y del código fuente". Todos los elementos, en fin, que permitan conocer las coordenadas principales del sistema de

---

131 Bien es verdad que, en sede de otras Administraciones, la autorización de la automatización la adopta un órgano concreto, particularmente competente en materia tecnológica, a petición del órgano competente para adoptar el acto a automatizar. Así, en el caso del Ayuntamiento de Madrid, según precisa en el Acuerdo de 18 de noviembre de 2021 de la Junta de Gobierno de la Ciudad de Madrid por el que se aprueban las directrices sobre actuación administrativa automatizada en el Ayuntamiento de Madrid y se modifica el Acuerdo de 5 de septiembre de 2019, de organización y competencias de la Coordinación General de la Alcaldía (BOAM nº 9020, de 22 de noviembre de 2021). Como también se ha dado el caso de que una norma con rango de ley implante actuaciones administrativas automatizadas, tal y como recuerda E. GAMERO CASADO (2023: 408).

132 La propuesta original de A. BOIX PALOP (2020) fue contraargumentada por autores como A. HUERGO (2020) y L. ARROYO (2020). El propio A. BOIX PALOP (2022: 90-105) ofrece sucesivamente una aproximación más matizada. Una aproximación general sobre la cuestión en BERNING PRIETO (2023: 95-130).

IA utilizado y dar muestras de su fiabilidad, ya que su mera enunciación implica que, con carácter general, deben estar contemplados.

La determinación de todos estos extremos no corresponderá, necesariamente, al mismo órgano al que se deba atribuir la actuación que aplica el sistema de IA[133]. Y en su fijación, no será necesario, como regla, asegurar la participación de los ciudadanos, inoperante por concepto en el grueso de los supuestos, por la complejidad de los sistemas y los términos en los que se configuran: típicamente por terceros a los que la Administración adquiere el sistema como producto o servicio. Por las razones que anticipamos en el primer apartado de esta ponencia, al describir someramente las características de los sistemas, y las que desarrollaremos coherentemente en el siguiente epígrafe, no tiene sentido contemplar semejantes trámites dentro de lo que no deja de ser la determinación de un elemento instrumental al ejercicio de la competencia conforme al procedimiento legalmente establecido, sin que el principio de buena administración ofrezca argumento añadido alguno para razonar en sentido distinto[134].

---

133 Como es sabido, el art. 45 de la Ley 30/1992 preveía que "los programas y aplicaciones electrónicos, informáticos y telemáticos que vayan a ser utilizados por las Administraciones Públicas para el ejercicio de sus potestades, habrán de ser previamente aprobados por el órgano competente, quien deberá difundir públicamente sus características" (art. 45). Asimismo, el art. 96.4 de la Ley 58/2003, de 17 de diciembre, General Tributaria señala que "los programas y aplicaciones electrónicos, informáticos y telemáticos que vayan a ser utilizados por la Administración tributaria para el ejercicio de sus potestades *habrán de ser previamente aprobados por ésta en la forma que se determine reglamentariamente*". En 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, en su 42 prescribe, en términos que recuerdan a los del 41 LRJSP, que será el Comité técnico estatal de la Administración judicial electrónica el órgano que, en caso de actuación automatizada, defina "las especificaciones, programación, mantenimiento, supervisión y control de calidad y, en su caso la auditoría del sistema de información y de su código fuente".

134 J. PONCE (2018) reclama que, en aplicación del principio de procedimiento debido, derivado en último término del principio de buena administración, se asegure un trámite de información pública y, en su caso, de audiencia, para la aprobación de los algoritmos y los códigos fuente subyacentes a los sistemas de Inteligencia Artificial, incluso cuando se elaboren por contratistas, que en este caso, según defiende, vendrían a ejercer funciones públicas como "entidades colaboradoras informáticas de la Administración". La simple lectura del art. 41 de la Carta de Derechos Fundamentales de la UE, que consagra el tantas veces evocado últimamente principio de buena administración, permite concluir que no hay en él garantía alguna distinta a algunas bien asentadas en el ordenamiento español, como bien pone de manifiesto G. FERNÁNDEZ FARRERES (2023), sin que entre ellas ninguna tenga proyección sobre lo que ahora nos ocupa.

La Evaluación de Impacto del sistema, cuya exigencia extensiva hemos defendido a la vista de las implicaciones que éste pueda tener, particularmente cuando implique la toma de decisiones, será el cauce natural en el que valorar todos estos extremos, incluyendo –en su caso- el mentado trámite de información pública, tal y como tuvimos ocasión de anticipar, máxime si integra una Evaluación de Impacto *ex* art. 35 RGPD[135]. Y ello en el bien entendido de que –conforme a la lógica del ciclo de vida del sistema- la clave estará en los controles operativos que se exijan durante la aplicación del sistema, particularmente los que se derivan de los requisitos que venimos describiendo exige la propuesta de RIA para los sistemas de alto riesgo, que convendría aplicar, en vía de principio, a la utilización de cualquier sistema de IA por Administraciones Públicas, aun en los casos en los que el propio sistema, como producto, no sea considerado de alto riesgo en el seno de dicha norma.

Estos controles permiten, hasta donde sea posible, conocer el funcionamiento de los sistemas. En ello abundaremos en los epígrafes sucesivos. Siendo en todo caso posible impugnar la actuación de que se trate, tal y como da por supuesto, como por otra parte es obvio, el mismo art. 41.2 al precisar que se deberá determinar el órgano que debe ser considerado responsable a efectos de impugnación, no debiendo ser otro que el que correspondiera si el acto se hubiera adoptado de forma analógica[136]. En sede de esa impugnación podrá plantearse el escrutinio del funcionamiento del sistema.

Las exigencias de la transparencia se verían, en definitiva, colmadas en este primer estadio por el simple mecanismo de dar publicidad a la circunstancia de la utilización de un sistema de IA para producir actos dentro de un determinado procedimiento, precisando el alcance de cuáles sean y la identificación -por refrendo de la competencia analógica- del órgano al que se le imputan y -correlativamente- del órgano ante el que pueden impugnarse, incluyendo una somera descripción de las coordenadas técnicas del sistema, en los términos que veremos en el último epígrafe de este apartado. Aunque, llamativamente, los preceptos de referencia y, en particular, el art. 41 LRJSP, no exigen

---

135 M. WIERZBOWSKI, M. (Coord.) (2022: 22) defienden este trámite de información pública en el contexto de la evaluación de impacto a la que deban someterse determinados sistemas dirigidos a la toma de decisiones.

136 Así lo hace expreso, oportunamente, el art. 44.3 de la Ley catalana 26/2010, de 3 de agosto, según quedó transcrito en la nota 129.

expresamente, esta publicidad[137], sí lo hace el art. 11.1.i) del RAFESP, para exigir que en las sedes electrónicas de cada organismo del sector público estatal se publique una relación actualizada de las actuaciones administrativas automatizadas que se despliegan en su ámbito de actuación, incluyendo una descripción de sus elementos técnicos y organizativos más significativos, en coherencia con lo que, por otra parte, exige el art. 5.1 de la Ley 19/2013, de 9 de diciembre, de Transparencia, Acceso a la Información Pública y Buen Gobierno.

El art. 16.1.l) de la Ley 1/2022, de 13 de abril, de Transparencia y Buen Gobierno de la Comunitat Valenciana, hace expresa una obligación semejante, específicamente en relación con los sistemas de IA. Como también lo hace, aun en términos menos exhaustivos, el art. 11.2 del Real Decreto-ley 2/2023, de 8 de marzo, de medidas urgentes de impulso a la inteligencia artificial en Extremadura.

3.- Parámetros de transparencia en relación con el funcionamiento *ad intra* de los sistemas de IA para la propia Administración: entre lo deseable y lo posible

Cuando se trata de asegurar la transparencia del funcionamiento de un sistema de IA hay varios planos en juego que frecuentemente se confunden. Asumiendo el neologismo "explicabilidad", de uso común en este ámbito, la misma –entendida como inteligibilidad, comprensión, de la mecánica del sistema y de sus resultados- se predica tanto respecto del usuario –la Administración, en nuestro caso, que lo utiliza en sus funciones, con independencia de que ella misma pueda jugar el papel de proveedora-, como respecto del receptor último de los resultados del sistema de IA, que en el caso de la Administración son los ciudadanos afectados, con mayor o menor alcance formal, por las actuaciones de las Administraciones públicas que apliquen estos sistemas. En uno y otro plano, hay que distinguir la inteligibilidad para los profanos de la predicable respecto de los expertos, siendo así que incluso para estos algunos sistemas de IA pueden resultar, como nos consta, incomprensibles.

Tratamos en este momento de la inteligibilidad del sistema mismo en clave doméstica, en la medida en que las Administraciones deben poder comprender el funcionamiento de los

[137] Frente a lo que exigía el precedente art. 45 de la Ley 30/1992, según quedó transcrito en la nota 132.

sistemas de IA que operen, para poder manejarlos adecuadamente, asegurarse su corrección y, en su caso, reaccionar ante incorrecciones. Respecto de la explicabilidad de sus resultados para sus destinatarios o, incluso, de su funcionamiento si es que plantean dudas al respecto, particularmente en vía de recurso, trataremos en el epígrafe siguiente, conscientes de que los planos *ad intra* y *ad extra* están íntimamente conectados.

En una primera aproximación, la garantía de la inteligibilidad del funcionamiento de un sistema de IA para los usuarios se debe mover, evidentemente, en un plano no técnico, donde prime la asequibilidad de las explicaciones con vistas a facilitar el uso del sistema y a evitar los recelos que pueda producir. En ese plano se proyectan las exigencias que el art. 13 de la propuesta de RIA integra, en coherencia con su consideración de los sistemas de IA como un producto que debe ser "manejable" para sus usuarios. De ahí que la primera previsión del artículo sea exigir que los sistemas de IA de alto riesgo se diseñen y desarrollen de un modo que garantice "que funcionan con un nivel de transparencia *suficiente* para que los usuarios interpreten y usen correctamente su información de salida" y con "un tipo y un nivel de transparencia *adecuados* para que el usuario y el proveedor cumplan las obligaciones oportunas previstas en el capítulo 3 del presente título".

Explicabilidad desde el diseño con vistas a su manejo por el usuario, a cuyo fin los sistemas de alto riesgo irán acompañados del "manual de instrucciones de uso" correspondiente, que incluirá "la información concisa, completa, correcta y clara que sea pertinente, accesible y comprensible para los usuarios", incluyendo expresamente referencia a las "características, capacidades y limitaciones del funcionamiento del sistema de IA de alto riesgo" (ap. 3.b) y a las medidas de vigilancia humana previstas en el art. 14, incluidas las medidas técnicas establecidas para facilitar la interpretación de la información de salida de los sistemas de IA por parte de los usuarios (ap. 3.b). Entre las precisiones exigidas por el apartado 3.b) están las referencias al nivel de precisión, solidez y ciberseguridad mencionado en el artículo 15 con respecto al cual se haya probado y validado el sistema de IA de alto riesgo y que puede esperarse de este, así como las circunstancias conocidas o previsibles que podrían afectar al nivel de precisión, solidez y ciberseguridad esperado (subap. ii), cualquier circunstancia conocida o previsible, asociada a la utilización del sistema de IA de alto riesgo conforme a su finalidad prevista o a un uso indebido razonablemente previsible, que pueda dar lugar a riesgos para la salud

y la seguridad o los derechos fundamentales (subap. iii) y, cuando proceda, especificaciones relativas a los datos de entrada, o cualquier otra información pertinente en relación con los conjuntos de datos de entrenamiento, validación y prueba usados, teniendo en cuenta la finalidad prevista del sistema de IA (subap. v).

La referencia expresa a los datos manejados -para su uso o, en su momento, para su desarrollo- pone en escena, una vez más, que son los datos tanto o más que los algoritmos utilizados y, menos aún, el código fuente, los que determinan la comprensibilidad del sistema. Y ni siquiera bastará en todo caso con eso –algoritmos y datos-, entendidos en su condición *sistemática*, máxime si se trata de algoritmos de aprendizaje y, más aun, en los casos de redes neuronales. En tales casos, la transparencia requiere conocer, sí, tanto los algoritmos de entrenamiento como los datos de entrenamiento (a razón de millones), particularmente si se quiere detectar sesgos y actuar sobre ellos; pero, además, el diseño mismo de la red [número de capas (en casos reales, aproximadamente unas mil), número de neuronas de cada capa, parámetro asociado a cada conexión] y el algoritmo de funcionamiento, si bien este es el extremo menos relevante, ya que usualmente se trata de algoritmos conocidos que implican aplicar fórmulas matemáticas a cada neurona en función de las entradas que recibe.

Baste esta somera descripción para poner de manifiesto que el verdadero funcionamiento de algunos sistemas de IA solo es asequible, y hasta donde lo sea en ocasiones, a través de un conocimiento experto. Y, tal y como tuvimos ocasión de apuntar en el primer apartado de este estudio, ante las "cajas negras" ni siquiera los expertos pueden explicarse a sí mismos el funcionamiento del sistema, de modo que se desarrollan líneas de investigación y actuación con vistas a lograr la explicabilidad de sistemas opacos basadas principalmente en aplicar métodos deductivos proyectados sobre las inferencias inductivas[138].

Las exigencias de trazabilidad y de auditabilidad surgen, en paralelo, como mecanismos imprescindibles para asegurar la verosimilitud del correcto funcionamiento de los sistemas complejos dentro de su opacidad. La exigencia de este tipo de mecanismos e instrumentos parece más realista que las llamadas de la ENIA a la utilización de

138 E. GAMERO CASADO (2023: 402) asume la imposibilidad de conocer la lógica del sistema en algunos casos.

"algoritmos transparentes y explicables". De este modo se favorece, además, la fiabilidad de los sistemas frente a los destinatarios de sus resultados, los ciudadanos destinatarios de las actuaciones de las Administraciones públicas, de verdaderos actos, en su caso.

Desde la perspectiva de la Administración, lo determinante será asegurar un sistema ponderado de vigilancia humana que tome como referencia el art. 14 de la propuesta de RIA, más arriba analizado. Y no solo para poder reaccionar frente a posibles fallos del sistema, sino también para validar sus resultados, de forma metódica o reactiva, lo que deberá valorarse en función de la potestad que se ejerza en cada caso. Sobre ello volveremos en el último apartado de esta ponencia, al tratar de los mecanismos de rendición de cuentas, pero antes debemos completar estas reflexiones sobre la transparencia de los sistemas valorando cómo ésta debe proyectarse sobre los ciudadanos.

3.- Parámetros de transparencia en relación con el funcionamiento *ad extra* de los sistemas de IA desde el punto de vista del destinatario: entre lo posible y lo deseable

Desde la perspectiva *ad extra*, la explicabilidad de los sistemas de IA aplicados por Administraciones Públicas se manifiesta, en una primera aproximación, en el alcance con el que puede y debe trasladarse al destinatario de las actuaciones públicas que se apoyen en sistemas de IA cómo funcionan estos. Cuestión que se suscita con verdadera densidad si se refiere a actuaciones de las que, en su caso, puedan reconducirse al espacio del ya analizado art. 41 LRJSP, particularmente cuando se trate de verdaderos actos decisorios.

Este planteamiento adquiere una especial dimensión, como sabemos, cuando datos personales están en juego, en los términos en los arts. 13.2.f) y 14.2.g), y 15.1.h), en relación con el 22, RGPD, en los casos de elaboración de perfiles o adopción de decisiones automatizadas en los términos fijados por este último. En tales supuestos, como manifestación del derecho a ser informado o del derecho de acceso del afectado por los tratamientos de datos personales que implica el sistema en cualesquiera de sus fases, deberá recibir "*al menos en tales casos, información significativa sobre la lógica aplicada, así como la importancia y las consecuencias previstas de dicho tratamiento para el interesado*".

Ya tuvimos ocasión de destacar cómo la AEPD, al vincular estas exigencias con la "explicabilidad" del tratamiento mediante IA, advierte de que la mera "referencia técnica a la implementación del algoritmo puede ser opaco, confuso, e incluso conducir a la fatiga informativa". Expresiones muy elocuentes que ponen de manifiesto que las explicaciones, para ser "significativas", deben ser asequibles para el ciudadano, al que no se le puede presumir un conocimiento experto, sin perjuicio de que se incluya, cuando corresponda, referencia a las auditorías o certificaciones aplicadas, como argumento de robustez y verosimilitud del tratamiento[139]. Esta aproximación, que vale para lo que se refiere al sentido y alcance de los tratamientos de datos personales involucrados en los sistemas de IA –en esta base se fundamenta, obviamente, la posición de la AEPD-, es perfectamente extensible a la comprensibilidad del funcionamiento del propio sistema.

¿Cómo materializar estos requerimientos en aras de la explicabilidad de los sistemas?. Resulta interesante a estos efectos el planteamiento de la legislación francesa que reconoce al interesado un derecho a ser informado, a demanda, de los parámetros de la solución de IA en función de cómo ha sido aplicada al caso[140]. Sería deseable introducir una previsión semejante en nuestro ordenamiento que, con todo, en vía de principio, es independiente de la debida motivación del acto –sobre la que volveremos en el apartado

---

139 AEPD (2020: 24). Tal y como pusimos de manifiesto supra (epígrafe III.3.A), el alcance de esta información significativa, ceñida, en el ámbito de las funciones de la Agencia, a lo referido a la protección de datos personales, debe ir dirigido a permitir entender el comportamiento del tratamiento, incluyendo aspectos como el detalle de los datos empleados para la toma de decisión y la precisión de la importancia relativa otorgada a cada uno, la calidad de los datos de entrenamiento y los tipos de patrones utilizados, los perfilados realizados y sus implicaciones, la existencia o no de supervisión humana y, particularmente, la referencia a auditorías, especialmente sobre las posibles desviaciones de los resultados de las inferencias, así como las certificaciones realizadas al sistema de IA.

140 En los términos literales del art. 4 de la Loi nº 2016-1321 du 7 de octobre 2016 pour une République numérique, "*une décision individuelle prise sur le fondement d'un traitement algorithmique comporte une mention explicite en informant l'intéressé. Les règles définissant ce traitement ainsi que les principales caractéristiques de sa mise en ouvre sont communiquées par l'administration à l'intéressé s'il en fait la demande*", mientras que el art. 6 prescribe, para garantizar la publicidad del manejo de los sistemas, que "*les administrations publient en ligne les règles définissant les principaux traitements algorithmiques utilisés dans l'accomplissement de leurs missions lorsqu'ils fondent des décisions individuelles*". Aquel derecho encuentra desarrollo en el art. 311-3-1-2 del Décret du 16 mars 2017, "*l´administration communique à la personne faisant l'objet d'une décision individuelle prise sur le fondement d'un traitement algorithmique, à la demande de celle-ci, sous une forme intelligible et sous réserve de ne pas porter atteinte à des secrets protégés par la loi, les informations suivantes : 1° Le degré et le mode de contribution du traitement algorithmique à la prise de décision ; 2° Les données traitées et leurs sources ; 3° Les paramètres de traitement et, le cas échéant, leur pondération, appliqués à la situation de l'intéressé; 4° Les opérations effectuées par le traitement*".

siguiente-. Conviene insistir, en todo caso, en que en la medida en que se mueve en el plano de la comunicación con el interesado, deliberadamente debe descartar honduras técnicas[141].

Una capa más profunda de densidad en la explicación del sistema puede ser exigible, particularmente en el contexto de la impugnación del acto que sea resultado o se apoye en la aplicación de una solución de IA. Antecedentes hay en nuestro ordenamiento con desigual alcance, partiendo de la Ley 19/2013, de 29 de diciembre, de Transparencia, Acceso a la Información y Buen Gobierno, en sintonía con referentes de ordenamientos cercanos[142].

En el marco de los arts. 2.b), 13 y 19.1 de esta ley se ha llegado a reconocer como "información pública", para garantizar el acceso a la misma, el "código fuente" y el "algoritmo matemático implementado por el (…) programa informático" del sistema catalán de designación de los miembros de los tribunales correctores de la PAU, por parte de la Agencia catalana de transparencia en resolución de 21 de septiembre de 2016[143]. Pero en ocasiones se han acogido excepciones para la accesibilidad a estas informaciones en el marco del art. 14 de la Ley, y muy en particular la relativa a la seguridad pública (ap. 1.d) y la que se deriva de la debida protección de los derechos de propiedad

---

141 PONCE (2018) también conecta una obligación semejante con el principio de buena administración, sin que en su opinión sea necesaria intermediación alguna del legislador. Y llega a afirmar que, de ser incomprensible la explicación para un sujeto medio, cabría "alegar la imposibilidad de cumplir lo exigido (lo que haría la decisión administrativa imposible y/o irracional, art. 47 Ley 39/2015 para actos, art. 9.3 Constitución española en general) y solicita la anulación judicial de la decisión adoptada, así como, en su caso, responsabilidad patrimonial por los daños que la actividad incomprensible de la Administración hayan causado en el destinatario". Conviene precisar, sin embargo, que el acto en sí puede ser perfectamente comprensible, aunque no lo sea la explicación del funcionamiento del sistema de IA que lo produzca o en el que se apoye la correspondiente decisión.

142 Son de cita común el caso en el que la Comisión francesa *d'accès aux documents administratifs* obligó a la *Direction générale des finances publiques* a hacer público el código fuente del programa de ordenador usado para calcular o el impuesto sobre los ingresos de las personas físicas mediante dictamen 20144578, de 8 de enero de 2015, y a hacer accesibles los códigos fuente de tres programas que desarrollan modelos con datos económicos usados por el Ministerio de Economía, mediante dictamen 20180276, de 19 de abril de 2018. Y el del Tribunal Administrativo Regional de Lazio-Roma cuando declaró, por su parte, en sentencia de 22 de marzo de 2017 (nº 3769), que un algoritmo que servía para la toma de decisiones automatizadas sobre movilidad de docentes por la Administración pública italiana, era un acto administrativo digital para concluir que los ciudadanos tienen derecho de acceso al mismo bajo la legislación de acceso a la información.

143 En los términos de la resolución 200/2017, de 21 de septiembre de 2016, que estimó las reclamaciones acumuladas 123 y 124/2016.

intelectual e industrial (ap. 1.j), solo aplicable si la propia Administración no es titular de los derechos sobre el sistema. Este fue el caso en el asunto que resolvió la SAN de 30 de diciembre de 2021, que respaldó la denegación del acceso al algoritmo aplicado en el sistema de validación de las solicitudes de reconocimiento de la condición de usuario vulnerable a los efectos de obtener el bono social eléctrico que había decidido el Consejo de Transparencia y Buen Gobierno[144].

Es de recordar, en todo caso, que ni el algoritmo, ni menos aún, el código fuente, serán necesariamente reveladores por sí mismos respecto del funcionamiento del sistema, tal y como puso de manifiesto respecto de este último la misma resolución de la Comisión catalana de 21 de septiembre de 2016[145]. Si a eso se unen limitaciones derivadas de la propia protección de datos o de la protección de la propiedad industrial o intelectual[146], las auditorías pueden jugar, de nuevo, un papel relevante para comprobar el buen funcionamiento del sistema[147], sin perjuicio de la posibilidad de llevar a cabo periciales expertas en contextos litigiosos. Sobre ellas reflexionaremos a continuación al plantearnos, como apartado final, los distintos mecanismos dirigidos a la rendición de cuentas de los sistemas de IA empleados por Administraciones Públicas.

V.- Mecanismos de rendición de cuentas por la utilización de sistemas de IA por Administraciones Públicas

144 SAN de 30 de diciembre de 2021 (Roj: SAN 5863/2021), que desestimó el recurso interpuesto por la fundación CIVIO contra la resolución del Consejo de Transparencia y Buen Gobierno de 18 de febrero de 2021 (R/0701/2018).

145 Código fuente no necesariamente revelador: Consejo catalán 123/2016: Ponce https://www.elnotario.es/index.php/hemeroteca/revista-77/opinion/opinion/8382-codigo-fuente-algoritmos-y-fuentes-del-derecho

146 Es expresiva la advertencia de la SAN de referencia, cuyo FJ 4 *in fine* precisa que "no existe ninguna norma que imponga a la Administración el desarrollo de aplicaciones con fuentes abiertas ni la adquisición de software libre". Téngase en cuenta a este respecto, que los pliegos que disciplinen la contratación de un contrato de servicios pueden asumir la no cesión de los derechos de propiedad intelectual o industrial que correspondan, según contempla el art. 308.1 LCSP, por más que G. VESTRI (2021: 384-385) y A.D. BERNING (2023: 124-127) defiendan lo contrario. No puede darse por hecho, pues, que las Administraciones Públicas sean titulares de los derechos de propiedad intelectual derivados de los sistemas de IA que utilicen, por lo que en tales casos no podrá suscitarse la posibilidad que contempla el art. 157.2 LRJSP de declarar abiertas las aplicaciones de cuyos derechos de propiedad intelectual sean titulares "cuando de ello se derive una mayor transparencia en el funcionamiento de la Administración pública".

147 PONCE (2019) asume también esta opción.

El tercer y último mecanismo que contempla el art. 23.1 de la Ley 15/2022, que hemos asumido como referencia para sistematizar las garantías que deben acompañar la utilización de IA por las Administraciones públicas y, en particular, la utilización de algoritmos para la toma de decisiones, se refiere a la rendición de cuentas. Un término que puede resultar equívoco, pero que aún con todo permite identificar determinados instrumentos para la supervisión y control del manejo de sistemas de IA por Administraciones Públicas. Cuestión principal a este respecto es la que implica indagar sobre el alcance del control judicial posible sobre los actos adoptados utilizando inteligencia artificial, lo que obliga a enfrentar con carácter previo la duda de con qué alcance puede o debe permitirse, y con qué autonomía respecto del ser humano, que efectivamente se adopten actos mediante estos sistemas.

Aparte habrá de tratarse sobre los márgenes que se le atribuyen a la recién creada Agencia de Inteligencia Artificial para supervisar y controlar la utilización de IA por las Administraciones Públicas, particularmente en relación con las auditorías y las evaluaciones de impacto que se puedan ejecutar para posibilitar esa rendición de cuentas y, por ende, para asegurar la fiabilidad de los sistemas de IA cuando se utilizan para desenvolver actuaciones administrativas.

Siendo estas las cuestiones a tratar, este apartado opera, en fin, en más de un extremo, como recapitulación de los anteriores[148].

1.- El pleno alcance del control judicial de las resoluciones adoptadas con IA: al hombre lo que es del hombre y a la máquina lo que es de la máquina con el régimen del silencio administrativo como referente

En los epígrafes precedentes, nos hemos planteado las garantías formales respecto de la puesta en conocimiento de los interesados del manejo de sistemas de Inteligencia Artificial por las Administraciones Públicas y de los parámetros de su funcionamiento, particularmente bajo las coordenadas que ofrece el art. 41 LRJSP en relación con las actuaciones automatizadas. Sin embargo, una cuestión ha quedado hasta ahora soslayada:

[148] No tratamos, sin embargo, deliberadamente, las cuestiones relativas a la utilización de inteligencia artificial por Administraciones Públicas, ya que no hay especialidades significativas que merezcan ser explicadas. Nos remitimos al respecto a R. MARTÍNEZ GUTIÉRREZ (2023: *in totum*).

¿hay algún límite para la utilización de soluciones de IA en la adopción de actos decisorios por parte de Administraciones Públicas?. Dicho de otro modo, ¿hay algunas decisiones que no deban adoptarse aplicando sistemas de IA?[149]

Se puede afirmar que está extendida la conclusión de que resulta inapropiado que se lleguen a adoptar decisiones en ejercicio de potestades discrecionales utilizando sistemas de IA. A tales efectos, como es bien sabido, se ha llegado a defender que debe darse una reserva de humanidad que prohíba el ejercicio de potestades discrecionales a través de sistemas de IA, por cuanto solo los seres humanos disponen de la empatía necesaria para adoptar decisiones que impliquen una opción entre varias posibles[150]. Algunas normativas autonómicas –como la catalana, la andaluza y la aragonesa- acogen, de hecho, este criterio en relación con las actuaciones administrativas automatizadas, descartando que puedan adoptarse en el caso de que exijan juicios de valor o impliquen la aplicación de criterios subjetivos[151]. El legislador alemán, por su parte, ha excluido la posibilidad de adoptar actos a través de medios automáticos cuando el marco normativo aplicable contemple un margen de decisión[152].

Esta aproximación a la cuestión merece varias reflexiones.

En estos argumentos subyace una desconfianza hacia la máquina que contrasta con una confianza en el ser humano y en su forma de razonar. Y ello a pesar de que, como adelantamos en su momento, no hay caja negra más extrema que el cerebro humano. Desentrañar las razones últimas que han empujado a un ser humano a tomar una decisión no es en absoluto fácil, ni siquiera para el propio ser humano que decide, y sin que ello esté impregnado necesariamente de deliberadas tergiversaciones. El ser humano que ha de adoptar decisiones en aplicación de las normas con autoridad, con capacidad de

149 Desde una lógica conceptual, con la vista puesta en la protección de los derechos humanos, son reveladoras las reflexiones de F. GAMPER (2023).

150 En este sentido, particularmente J. PONCE (2018) y (2022) y, de forma más matizada, E. GAMERO CASADO (2023: 448-452).

151 Este es el caso del art. 44.2 de la Ley catalana 26/2010, de 3 de agosto, de régimen jurídico y de procedimiento de las administraciones públicas de Cataluña; del 40 del Decreto 622/2019, de 27 de diciembre, de administración electrónica, simplificación de procedimientos y racionalización organizativa de la Junta de Andalucía, y del art. 43 de la Ley 5/2021, de 29 de junio, de Organización y Régimen Jurídico del Sector Público Autonómico de Aragón.

152 Art. 35a de la Verwaltungsverfharengesetz.

imponer su criterio por tratarse del titular de un órgano administrativo, ha sido entrenado –como se entrena a un algoritmo- específicamente a través de una formación jurídica contrastada en unas pruebas *ad hoc* y con una formación específica posterior a su selección. Y a partir de ahí el ordenamiento confía en sus decisiones otorgándoles la presunción de validez, sin perjuicio de posibilitar que sean controladas, eso sí, por otro ser humano igualmente entrenado para adoptar decisiones en Derecho. Los razonamientos profundos conforme a los cuales adopten uno y otro sus decisiones son, incluso para ellos mismos, un enigma, sin que la motivación con la que se formulen los actos y las sentencias haga otra cosa que dar una explicación, con coordenadas jurídicas, de lo decidido. Así lo puso magistralmente de manifiesto el añorado Alejandro Nieto en su "El arbitrio judicial", elocuentemente subtitulado "Entrando en la mente del juez", con argumentos que se pueden predicar del propio humano titular del órgano administrativo cuyos actos se someten a control[153].

Para ese control, conviene advertirlo, poco importan las razones últimas que hayan llevado al titular del órgano administrativo a adoptar su decisión. Lo que importa es el acto en sí, y como tal ha de ser sometido a escrutinio para valorar su conformidad con el ordenamiento jurídico. Puestas así las cosas, que en la adopción del acto el ser humano sea sustituido por un sistema de Inteligencia Artificial, no resulta tan inasumible. Es más, incluso se puede afirmar que los sistemas de Inteligencia Artificial podrían llegar a asegurar soluciones más consistentes y uniformes, a lo que se suma la batería de garantías de transparencia del proceso decisorio que, a pesar de sus francas limitaciones, en las que hemos hecho hincapié, son incomparables con la condición impenetrable del razonamiento del ser humano, a pesar de los términos en que se exteriorice su explicación.

---

[153] A. NIETO (2021: 522) lo expresa descarnadamente: "Los juristas estamos obligados a imponer el rigor y a favorecer la seguridad, pero no podemos ignorar que esto es imposible en la vida real. El Derecho no es una ciencia exacta, ni se refiere siempre a hechos precisos ni mucho menos maneja teorías inequívocas y principios fiables. En el Derecho, las leyes y sus operadores (abogados, funcionarios y jueves) se ven constantemente influidos y desviados por imprevisibles circunstancias concretas, necesidades sociales antes desconocidas, presiones políticas torticeras, interferencias personales caprichosas y hasta vicios inconfesables que convierten el Derecho en un universo en el que la precisión y la previsión son imposibles. No hay respuestas seguras para todas las dudas y conflictos. Un cambio político, un giro doctrinal o la imaginación de un tribunal pueden enviar al cesto de los papeles bibliotecas enteras con sus sesudos dictámenes y cientos de sentencias que habían estado corriendo durante años como verdadero Derecho".

Otra cosa es qué capacidades reales ofrezca la Inteligencia Artificial en su estadio actual, en función de la tipología de algoritmos y de su fiabilidad[154]. En esa posición *realista* parece haberse colocado la propuesta de RIA cuando identifica como sistemas de alto riesgo los "*destinados a ayudar a una autoridad judicial en la investigación e interpretación de hechos y de la ley, así como en la aplicación de la ley a un conjunto concreto de hechos*", sin concebir que los sistemas puedan ser directamente decisorios.

Si bajo coordenadas de eficacia y eficiencia se plantea la posibilidad de que sistemas de IA se adentren en la toma de decisiones, la decisión al respecto puede encontrar un valioso punto de referencia, más allá de la distinción entre potestades regladas y discrecionales, en el tratamiento del silencio administrativo. Piénsese que, de hecho, la institución del silencio supone reconocer de forma *automática*, por el mero transcurso del tiempo, un efecto jurídico, que en los supuestos del silencio positivo es estimatorio de lo originalmente solicitado, sin que la resolución extemporánea pueda tener otro sentido que el confirmatorio de lo que es un verdadero acto *ex* art. 24.2 y 3, siempre a salvo la posibilidad de la revisión de oficio.

Este mismo referente puede servir para delimitar los márgenes de acción de la supervisión humana que hemos defendido necesaria en relación con la aplicación de sistemas de IA por las Administraciones Públicas, más allá de los casos en los que venga exigida por su consideración como sistema de alto riesgo en el marco de la propuesta de RIA. En los supuestos en los que se pudiera llegar a decidir integrar una solución de IA para la toma de decisiones, es imprescindible valorar en qué grado el supervisor humano debe validar el resultado de la máquina: en sentido activo –solo habrá decisión cuando el humano valide a la máquina- o pasivo –el humano puede descartar puntualmente la decisión de la máquina-[155]. Si bien en el primer caso no habría propiamente toma de decisión por el

---

154 I. MARTÍN DELGADO (2023: 178-180), por su parte, de forma ponderada, justifica la exclusión del uso de soluciones de IA para el ejercicio de potestades discrecionales en el estadio actual de la tecnología. A. HUERGO (2022: 83-85) y G. VESTRI (2021: 370-380) descartan que los algoritmos predictivos puedan sustentar por sí mismos la toma de decisiones.

155 Puede incluso plantearse en función de si la información de salida del sistema de IA es desestimatoria o estimatoria, siendo así que la validación humana solo se requiera en el primer supuesto, como es el caso del Sistema Europeo de Información y Autorización de Viajes, referido por O. MIR PUIGPELAT (2023: 692-703). A un planteamiento similar apela E. GAMERO CASADO (2023: 451-452) en sintonía con su consideración de las limitaciones empáticas de la IA. Se debe poner de manifiesto, en todo caso, que la defensa de los intereses generales impide

sistema, en ambos supuestos hay que evitar el consabido riesgo de automatismo, sin perjuicio de exigir una motivación exhaustiva de la decisión de separarse del resultado arrojado por el sistema para evitar que el humano distorsione la lógica del mismo seleccionado libérrimamente los supuestos en los que la acoge y los que no. Se trata, en último término, de depurar los supuestos en los que el sistema desemboque en "soluciones absurdas", contrarias como tales al ordenamiento jurídico[156].

Sea como fuere, el acto que sea fruto, en mayor o menor grado, de un sistema de IA está –evidentemente- sometido a control jurisdiccional. Sobre los parámetros en los que este sea posible es oportuno hacer algunas consideraciones añadidas.

Si nos movemos en la órbita del art. 41 LRJSP, el incumplimiento de los requisitos procedimentales que establece para la implantación de sistemas de IA que automatizaran la toma de decisiones determinaría por sí mismo una causa de anulabilidad de los actos adoptados en su aplicación. Si los incumplimientos procedimentales determinaran una vulneración de la legislación de protección de datos, el vicio sería determinante de nulidad por afectación al derecho fundamental del art. 18.4 CE en función de cuál pueda ser el extremo vulnerado, desde la falta de base legal para el tratamiento, hasta el incumplimiento de los requisitos del art. 22 RGPD, pasando, en su caso, por los incumplimientos del Esquema Nacional de Seguridad o la no elaboración de la Evaluación de Impacto cuando fuera exigible en los términos del art. 35[157]. Del mismo modo, los incumplimientos podrían determinar un vicio de nulidad si implicasen la generación de sesgos discriminatorios en el contexto del art. 14 CE, para cuya detección

asumir este criterio de forma indiscriminada, por cuanto hay decisiones estimatorias de según de qué solicitudes de un particular –pongamos por ejemplo una que afecte al medioambiente- para las que puede resultar crítica la supervisión humana. La toma en consideración de los criterios para la delimitación de los supuestos de silencio positivo o negativo se demuestra, también aquí, útil.

156 E. GARCÍA DE ENTERRÍA (1997: 414-415) y (2000: 381) apela a un principio interpretativo general que "prohíbe llegar a soluciones absurdas", según destaca G. FERNÁNDEZ FARRERES (2023).

157 J. VALERO TORRIJOS (2023: 388-391) identifica, en particular, la vulneración de las limitaciones establecidas en el art. 5 RGPD; el no alineamiento con el Esquema Nacional de Seguridad; la carencia de base jurídica adecuada para el tratamiento de los datos en función de su naturaleza y alcance o por falta de competencia de la Administración actuante; el incumplimiento de los requisitos del art. 22 RGPD cuando resultara de aplicación y la no ejecución de la evaluación de impacto cuando fuera obligatoria.

y depuración la propia Ley 15/2022 introduce, como pusimos de manifiesto, mecanismos cuya virtualidad está por ver.

La comprobación de unos y otros extremos puede resultar relativamente fácil a partir del expediente, en el que han de constar pormenorizadamente los elementos del sistema aplicado, a los que se les deberá dar acceso bajo las coordenadas que describimos en el apartado anterior. Más dificultad parece que podría plantear controlar el fondo de los actos, particularmente los que limiten derechos e intereses legítimos y los discrecionales, en relación con lo cual se ha venido suscitando la cuestión del debido alcance de su motivación, en los términos en que viene exigida por el art. 35.1, particularmente en su apartado a) y i) LRJPAC[158].

Repárese en que el precepto requiere a tal fin "una sucinta referencia de hechos y fundamentos de Derecho". Y repárese también en que el 88.6 asume que la aceptación de informes o dictámenes servirá de motivación a la resolución cuando se incorpore a la misma, lo cual da especial cobertura a los supuestos de supervisión humana activa, asumiendo que el informe o dictamen ofrecido por el sistema de IA deberá ofrecer una motivación. Lo relevante será que se revelen las coordenadas que pongan de manifiesto que la decisión adoptada es (pretendidamente) conforme con el ordenamiento jurídico en términos suficientemente elocuentes como para valorar la oportunidad de recurrirla –no se olvide que el requisito de motivación es medio para la garantía de la tutela judicial efectiva-. De modo que para ello, cuando se utilicen sistemas de IA, parece que no ha de bastar con poner de manifiesto los antecedentes de hecho y los parámetros jurídicos aplicados al caso, sino que se hace igualmente necesario exteriorizar los parámetros básicos de funcionamiento del sistema en términos que hagan comprensibles los mecanismos determinantes de la decisión. En estos términos se ha pronunciado el TJUE en su sentencia de 21 de junio de 2022, dictada en el asunto *Ligue de droits humains*[159],

---

158 E. GAMERO CASADO (430-432) sintetiza la doctrina sobre la cuestión, poniendo de manifiesto que mayoritariamente ha abordado la cuestión desde la perspectiva de la transparencia.

159 Dictado en el asunto C-817/19, en el que se resolvió una cuestión prejudicial planteada por el Tribunal Constitucional belga acerca del régimen europeo de tratamiento de los datos del registro de nombres de pasajeros. En los términos literales de sus apartados 210 y 211, "210.- En particular, las autoridades competentes deben asegurarse de que el interesado, al que no necesariamente se permite, en el procedimiento administrativo, tomar conocimiento de los criterios de evaluación predeterminados y de los programas que aplican tales criterios, pueda comprender el funcionamiento de esos criterios y de esos programas, de forma que pueda decidir,

con argumentos similares a los que ya se habían anticipado en algunos casos de referencia del Derecho comparado, como las resoluciones del Consejo de Estado italiano de 8 de abril y 13 de diciembre de 2019[160].

Lo que es claro en todo caso es que el acto que sea fruto, en mayor o menor medida, de un sistema de IA está sometido al escrutinio del juez o tribunal contencioso, debiendo evitarse también en este escenario el riesgo de automatismo. El juez o tribunal debe valorar la conformidad a Derecho de una decisión que no es infalible por el hecho de haber sido adoptada por o con apoyo de la máquina. En esta valoración podrá, a su vez, apoyarse en soluciones de IA, según contempla el RIA[161]. Pero es al que el ordenamiento le otorga la capacidad de decidir, cerrando el ciclo el último órgano jurisdiccional ante el que quepa recurso. En esta paradoja circular se desenvuelve el reto que supone introducir sistemas de IA en la adopción de decisiones jurídicas.

---

con pleno conocimiento de causa, si ejerce o no su derecho a un recurso judicial, garantizado en el artículo 13, apartado 1, de la Directiva PNR, con objeto de cuestionar, en su caso, el carácter ilícito y, en particular, discriminatorio de los mencionados criterios (véase, por analogía, la sentencia de 24 de noviembre de 2020, Minister van Buitenlandse Zaken, C-225/19 y C-226/19, EU:C:2020:951, apartado 43 y jurisprudencia citada). Así debe ocurrir también con los criterios de revisión mencionados en el apartado 206 de la presente sentencia. 211.- Por último, al sustanciar un recurso interpuesto con arreglo al artículo 13, apartado 1, de la Directiva PNR, el juez que ejerce el control de la legalidad de la resolución adoptada por las autoridades competentes y, salvo en los casos de amenazas para la seguridad del Estado, el propio interesado, deben poder acceder al conjunto de motivos y pruebas sobre cuya base se ha adoptado esta resolución (véase, por analogía, la sentencia de 4 de junio de 2013, ZZ, C-300/11, EU:C:2013:363, apartados 54 a 59), incluidos los criterios de evaluación predeterminados y el funcionamiento de los programas que aplican estos criterios".

160 Asuntos núm. 2270 y 8472, respectivamente. I. MARTÍN DELGADO (2023: 162-172) se decanta por exigir que, más allá de las razones que se ofrezcan para el caso concreto, se acredite que el sistema funciona correctamente, en conexión con una obligación de publicidad activa.

161 El caducado Proyecto de Ley de Eficiencia Digital del Servicio Público de Justicia la contemplaba igualmente, pero limitada a la labores meramente auxiliares. Conforme a su art. 57, "1. Se considera actuación asistida aquella para la que el sistema de información de la Administración de Justicia genera un borrador total o parcial de documento complejo basado en datos, que puede ser producido por algoritmos, y puede constituir fundamento o apoyo de una resolución judicial o procesal. 2. En ningún caso el borrador documental así generado constituirá por sí una resolución judicial o procesal, sin validación de la autoridad competente. Los sistemas de la Administración de Justicia asegurarán que el borrador documental sólo se genere a voluntad del usuario y pueda ser libre y enteramente modificado por éste". Sobre las claves del fenómeno, no solo en la toma de decisiones, S. BARONA VILAR (2022: *in totum*).

2.- Mecanismos de supervisión y control: en particular, auditorías y evaluaciones de impacto. Luces y sombras en el papel de la Agencia de Supervisión de la Inteligencia Artificial *avant la lettre*

Tal y como hemos tenido ocasión de poner de manifiesto en epígrafes previos de este estudio, los sistemas de IA, regulados desde su consideración como productos potencialmente generadores de riesgos a lo largo de su ciclo de vida, reclaman con naturalidad mecanismos de seguimiento de su funcionamiento que permitan detectar los errores o efectos no deseados, para reaccionar consecuentemente. En este planteamiento, conscientes de que la complejidad de los sistemas de IA, máxime los de mayor grado de autonomía, impide asegurar un riesgo 0, los mecanismos de seguimiento cumplen por sí mismos la función descrita e indirectamente son argumentos para generar la imprescindible confianza en los mismos.

En estos términos, instrumentos como las auditorías y evaluaciones de impacto se presentan como mecanismos perfectamente asentados en la regulación de gestión de riesgos, con manifestaciones muy claras –como tuvimos ocasión de describir- en la esfera de la protección de datos personales. Su razón de ser en relación con los sistemas de Inteligencia Artificial utilizados en la órbita de las Administraciones Públicas trasciende, sin embargo, la protección de datos, de modo que hay buenas razones para postular la oportunidad de introducir mecanismos de este tipo para asegurar la rendición de cuentas de las soluciones de IA en ese escenario.

A falta de una previsión más perfilada al respecto, no debemos olvidar que el art. 23 de la Ley 15/2022 que asumimos como norma de cabecera para sistematizar los mecanismos de garantía de la utilización de sistemas de IA en relación con la toma de decisiones por las Administraciones Públicas asume que "se promoverá la realización de evaluaciones de impacto que determinen el posible sesgo discriminatorio", en expresión perfectamente extrapolable a otros ámbitos de riesgo. El art. 41LRJSP, por su parte, contempla como uno de los elementos a especificar cuando se implante una actuación administrativa automatizada, la "supervisión y control de calidad y, en su caso, auditoría del sistema de información y del código fuente". Se trata, en definitiva, de garantizar mecanismos de seguimiento y supervisión del funcionamiento de los sistemas que soportan la actuación

automatizada, en unos términos que son perfectamente predicables en los supuestos en los que se apliquen a tal efecto sistemas de Inteligencia Artificial.

Las razones para la auditabilidad y evaluación de los sistemas IA aplicados por Administraciones públicas se proyectan, aun con distintos grados de intensidad, en relación con cualesquiera de sus usos, aun cuando resulten críticos en la órbita de la toma de decisiones a través o con el apoyo de sistemas de IA. Así parece asumirlo, de hecho, el Real Decreto 729/2023, de 22 de agosto, por el que se aprueba el Estatuto de la ya mencionada Agencia Española de Supervisión de Inteligencia Artificial, creada como organismo dirigido a asegurar la más plena aplicación de la legislación europea sobre Inteligencia Artificial, cumpliendo el papel del supervisor nacional que la propuesta de RIA contempla, a pesar de que este instrumento normativo está todavía en proceso de aprobación[162].

El art. 25.b).4º del Estatuto atribuye, en efecto, al Departamento de Sistemas de Inteligencia Artificial orientados a las Administraciones Públicas, al que ya hemos tenido ocasión de hacer mención, la función de "emitir informes sobre el impacto generado por un sistema de inteligencia artificial puesto en marcha, utilizado o que se esté desarrollando por las diferentes entidades del Sector Público, AP (app) o puestos en marcha desde otro u otros departamentos ministeriales", precisando que, "en concreto, se realizarán evaluaciones de impacto para que los algoritmos involucrados en la toma de decisiones que se utilicen en las administraciones públicas tengan en cuenta criterios de minimización de sesgos, transparencia y rendición de cuentas", en expresión que viene a refrendar la contenida en el reiteradamente citado art. 23 de la Ley 15/2022.

---

[162] La creación de la Agencia fue prevista por la Disposición adicional centésimo trigésima de la Ley 22/2021, de 28 de diciembre, de Presupuestos Generales del Estado para el año 2022, siendo efectivamente creada por la Ley 28/2022, de 21 de diciembre, de fomento del ecosistema de las empresas emergentes, en cumplimiento de la exigencia contenida en el art. 91 LRJSP. En los términos del apartado 2 de aquella, "esta Agencia actuará con plena independencia orgánica y funcional de las Administraciones Públicas, de forma objetiva, transparente e imparcial, llevando a cabo medidas destinadas a la minimización de riesgos significativos sobre la seguridad y salud de las personas, así como sobre sus derechos fundamentales, que puedan derivarse del uso de sistemas de inteligencia artificial. Estas medidas incluirán actuaciones propias, actuaciones en coordinación con otras autoridades competentes, cuando sea aplicable, y actuaciones de apoyo a entidades privadas".

El mismo Departamento, desde una perspectiva más general, tiene atribuida la función de "coordinar, como organismo de referencia a nivel estatal en materia de inteligencia artificial, al resto de departamentos ministeriales y administraciones públicas para garantizar el cumplimiento de los requisitos exigidos en la normativa nacional o europea en materia de inteligencia artificial" [art. 25.b).3º]. El Departamento de certificación, instrucción y supervisión, por su parte, integrado como está en la Subdirección de Certificación, Evaluación de Tendencias, Coordinación y Formación en Inteligencia Artificial (sic), integra entre sus funciones la de supervisar de oficio los sistemas de inteligencia artificial utilizados por las administraciones públicas, así como la emisión de informes vinculantes y actas que decida sobre la continuidad de dichos sistemas y/o su puesta en marcha [art. 26.a).6º], como función específicamente relacionada con el manejo de sistemas de inteligencia artificial por Administraciones públicas frente a la más genérica función de llevar a cabo la labor de instrucción de los expedientes sancionadores, la supervisión de los sistemas de inteligencia artificial y, en su caso, la propuesta de medidas correctoras para garantizar el cumplimiento de la normativa reguladora en el ámbito de la inteligencia artificial [art. 26.a).5º].

La Agencia de Supervisión está llamada, pues, a jugar un papel protagonista en la evaluación de la continuidad de los sistemas de IA utilizados por las Administraciones Públicas. Para ejercer esta función, será de radical importancia la facultad de acceso extensivo que el art. 64 de la propuesta de RIA le confiere respecto a los datos manejados y documentación de funcionamiento de los sistemas de alto riesgo, incluyendo, en caso necesario el acceso al código fuente del sistema. Facultad que se reconoce igualmente -aunque en una extensión más limitada y, en su caso, a través de la propia Agencia- a las autoridades encargadas de la protección de los derechos fundamentales potencialmente afectados por los sistemas, y siempre respetando las obligaciones de confidencialidad que impone el art. 70, particularmente atentas a la protección de los derechos de propiedad intelectual y la información empresarial confidencial o los secretos comerciales, incluido el código fuente.

El despliegue de semejantes funciones por parte de la Agencia plantea, sin embargo, no pocos interrogantes de principio, a los que se suman múltiples incertidumbres respecto de cada una de las funciones atribuidas. Plantea, en primer lugar, severas dudas que la

Agencia respete las exigencias de independencia que impone el art. 59.1 de la propuesta de RIA, toda vez que su Consejo Rector integra, en los términos del art. 15 de su Estatuto, sustancialmente a miembros de la estructura ministerial con rango de subdirector general, todos ellos nombrados por el titular del Ministerio de Asuntos Económicos y Transformación Digital. Esta composición añade dudas a la verosimilitud, desde un punto de vista jurídico, de la apuesta de someter al escrutinio de la Agencia las decisiones de implantar o continuar sistemas de IA por parte de todas las Administraciones Públicas. Labor, por otra parte, difícilmente articulable en la práctica si la IA llega a adquirir en el seno de actividad pública la expansión que desde el propio Gobierno, en el marco de la ENIA, se le pretende.

En lo que hace a las auditorías, para las que es fundamental que se cumplan los mecanismos de registro y seguimiento de funcionamiento de los sistemas del tipo de los que, como vimos, exige el RIA para los sistemas de alto riesgo y la propila legislación de protección de datos, resultará en ocasiones oportuno designar grupos de expertos independientes[163], entre los que –como con buen criterio se ha defendido- debe haber juristas[164]. Las auditorías podrán ser, en todo caso, internas o externas, automáticas o deliberadas, todo ello en función del objetivo y fin con el que se diseñen. De asegurar confianza en los sistemas, de forma proporcionada a los riesgos que generen, se trata.

V.- Unas breves reflexiones para concluir

Las múltiples y heterogéneas aplicaciones que se engloban en el concepto Inteligencia Artificial abren a las Administraciones Públicas un conjunto de posibilidades para la consecución de los intereses generales que, sin embargo, no están exentas de incertidumbres y riesgos. Su utilización en las actuaciones públicas tiene que venir revestida de suficientes garantías, pues solo así resultará verosímil y sostenible como medio, y no fin, que es.

---

163 M. WIERZBOWSKI (Coord.) (2022: 21-22) aboga por esta fórmula en el caso de los sistemas que se consideren de alto riesgo.

164 En particular, E. GAMERO CASADO (2021b) y (2023: 416-421) ha venido defendiendo la presencia de expertos en Derecho en la ejecución de la supervisión humana de los sistemas.

Si partimos de este carácter, por concepto, instrumental, de la Inteligencia Artificial, la fascinación que produce debe atemperarse. Si es dudoso que la Inteligencia Artificial pueda llegar a transformar los parámetros del Derecho Administrativo, sin duda en el estadio tecnológico actual está lejos de ser así. Su ordenación responde, desde el punto de vista regulatorio, a los parámetros típicos de la regulación de riesgos, y a todos sus instrumentos acude la propuesta de Reglamento europeo, de forma por momentos solapada con la aproximación que ofrece el Reglamento General de Protección de Datos, de aplicación extensiva en unas tecnologías que encuentran en los datos, tanto o más que en los tan traídos y llevados algoritmos, el elemento clave de su desarrollo.

Análisis y gestión de riesgos, Evaluaciones de impacto, Certificaciones, Auditorías, Sellos de calidad… se colocan en escena como mecanismos necesarios para disciplinar, que no eliminar, unos riesgos que generan desconfianza y, por ende, temor, en lo que no deja de regularse como un producto cuya libre circulación debe garantizarse. Cuando este producto se integra en el desenvolvimiento de actuaciones públicas, los parámetros del Derecho administrativo más clásico se reafirman. Incluso cuando se plantea la adopción de actos a través de soluciones de Inteligencia Artificial, sometidos a control judicial sin fisuras.

En el estadio actual de la tecnología, la supervisión humana se hace imprescindible en los sistemas de alto riesgo. La máquina no puede sustituir plenamente al humano que encarna el órgano administrativo actuante. Son sus carencias de racionalidad las que deben identificar la imposibilidad de aplicar soluciones de Inteligencia Artificial a según qué esferas de la actuación pública, particularmente las de carácter decisorio, sin que sea necesariamente la consideración del ejercicio de potestades discrecionales la que marque el criterio.

Son estos criterios basados en las capacidades mismas de los sistemas los que deben guiar las exclusiones de uso de la Inteligencia Artificial, más que los temores que infunde la falta de inteligibilidad plena de su funcionamiento. La mente del ser humano sigue siendo tierra ignota para los propios neurólogos, sin que haya mecanismo alguno para rastrear su funcionamiento. Frente a ello la introducción de los mecanismos de seguimiento, supervisión y control que, según hemos descrito en esta ponencia, disciplinan el uso de

sistemas de Inteligencia Artificial por las Administraciones Públicas, ofrecen –aun con sus debilidades- garantías incomparables para embridar el modo en que se producen los actos y, correlativamente, permitir su control.

Esta reflexión no impide terminar estas páginas con una nueva llamada a la prudencia en el uso de soluciones de Inteligencia Artificial por parte de las Administraciones Públicas. Las modas no son buenas consejeras y toda decisión al respecto debe estar guiada por criterios sopesados de eficacia y eficiencia que tengan en cuenta, muy en particular, que el riesgo 0 no existe para los derechos fundamentales e intereses públicos y privados en escena.

BIBLIOGRAFÍA

-ARROYO JIMÉNEZ, L. (2020), Algoritmos y reglamentos, Almacén del Derecho, entrada de 25 de febrero de 2020, disponible en https://almacendederecho.org/algoritmos-y-reglamentos
-BARONA VILAR, S. (2022), La seductora algoritmización de la justicia. Hacia una justicia posthumanista (Justicia+) ¿utópica o distópica?, El Cronista del Estado Social y Democrático de Derecho, nº 100.
-BERNING PRIETO, A. D. (2023), La naturaleza jurídica de los algoritmos, en en GAMERO CASADO, E. (Dir.), Inteligencia artificial y sector público. Retos, límites y medios, Tirant Lo Blanch, Valencia.
-BOIX PALOP, A. (2020), Los algoritmos son reglamentos: la necesidad de extender las garantías propias de las normas reglamentarias a los programas empleados por la Admnistración para la adopción de decisiones", Revista de Derecho Público: Teoría y Método, nº 1.

(2022), Transparencia en la utilización de inteligencia artificial por parte de la Administración, Cronista del Estado social y democrático de Derecho, nº 100.
-COTINO HUESO, L. (2021), Hacia la transparencia 4.0: el uso de la inteligencia artificial y el big data para la lucha contra el fraude y la corrupción y las (muchas) exigencias constitucionales, en RAMIO, C. (Coord.), Repensando la Administración digital y la innovación pública, INAP, Madrid, pp. 169-196.

(2023), Discriminación, sesgos e igualdad de la inteligencia artificial en el sector público, en GAMERO CASADO, E. (Dir.), Inteligencia artificial y sector público. Retos, límites y medios, Tirant Lo Blanch, Valencia.
- CRIADO, J.I. (2021), Inteligencia Artificial (y Administración Pública), Eunomía. Revista en Cultura de la Legalidad, 20, pp. 348-372.
-ESTEVE PARDO, J. (2023), Principios de Derecho regulatorio. Servicios económicos de interés general y regulación de riesgos, 2ª ed., Marcial Pons, Madrid.
-ESTEVEZ ALMENZAR, M. et al (2022), Glossary of human-centric artificial intelligence, JRC, Unión Europea, Luxembourg.
-FERNANDEZ FARRERES, G. (2023), El principio de buena administración según la doctrina de la Sala Tercera del Tribunal Supremo, Revista Española de Derecho Administrativo (en prensa).
-GAMERO CASADO, E. (2021a), El enfoque europeo de Inteligencia Artificial, Revista de Derecho Administrativo, CDA 20.

(2021b), *Compliance* (o cumplimiento normativo de desarrollos de Inteligencia Artificial para la toma de decisiones administrativas, Diario la Ley, nº 50.

(2023), Las garantías de régimen jurídico del sector público y del procedimiento administrativo común frente a la acticiada automatizada y la inteligencia artificial, en GAMERO CASADO, E. (Dir.), Inteligencia artificial y sector público. Retos, límites y medios, Tirant Lo Blanch, Valencia.
-GAMPER, F. (2023), The Limits of AI Decision-Making. Are There Decisions Artificial Intelligence Should Not Make?, en QUINTAVILLA, A. y TEMPERMAN, J. (Eds.), Artificial Intelligence and Human Rights, Oxford University Press.
- GARCÍA DE ENTERRÍA, E. (1997), Una reflexión sobre la supletoriedad del Derecho del Estado respecto del de las Comunidades Autónomas (sentencias constitucionales 118/1996, de 27 de junio, y 61/1997, de 20 de marzo), Revista Española de Derecho Administrativo, nº 95.

(2000), Sobre la ejecutoriedad inmediata de las medidas cautelares recurridas en casación. Una explicación rectificativa", Revista de Administración Pública, nº 153.

-GÓMEZ, E.; HUPONT TORRES, I.; SÁNCHEZ, I., y FERNÁNDEZ LLORCA, D. Unión Europea: una perspectiva científico-técnica, en GAMERO CASADO, E. (Dir.), Inteligencia artificial y sector público. Retos, límites y medios, Tirant Lo Blanch, Valencia.
-GONZÁLEZ CABANES, F. y DIAZ DIAZ, N. (2023), ¿Qué es la Inteligencia Artificial?, en GAMERO CASADO, E. (Dir.), Inteligencia artificial y sector público. Retos, límites y medios, Tirant Lo Blanch, Valencia.
-HAN, B.-Ch. (2021), No-cosas. Quiebras del mundo de hoy, Taurus, Madrid.
-HARARI, Y. N. (2016), Homo Deus. Breve historia del mañana, Debate, Barcelona.
- HUERGO LORA, A. (2020), Una aproximación a los algoritmos desde el Derecho administrativo, en HUERGO LORA, A. (Dir.) y DIAZ GONZÁLEZ, G. M. (Coord.), La regulación de los algoritmos, Aranzadi, Cizur Menor.
(2022), Gobernar con algoritmos, gobernar los algoritmos, Cronista del Estado Social y Democrático de Derecho, nº 100.
(2023), Hacia la regulación europea de la Inteligencia Artificial, en GAMERO CASADO, E. (Dir.), Inteligencia artificial y sector público. Retos, límites y medios, Tirant Lo Blanch, Valencia.
-LARSON, E.J. (2022), El mito de la Inteligencia Artificial, Schackleton Books.
- MANZONI, M., MEDAGLIA, R., TANGI, L., VAN NOORDT, C., VACCARI, L. y GATTWINKEL, D. (2022), AI WatchRoad to the adoption of Artificial Intelligence by the Public Sector: A Handbook for Policymakers, Public Administrations and Relevant Stakeholders, EUR 31054 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-52132-7, doi:10.2760/288757, JRC129100.
-MARTÍN DELGADO, I. (2009), Naturaleza, concepto y régimen jurídico de la actuación administrativa automatizada, en RAP, nº 180.
(2023), La aplicación del principio de transparencia a la actividad administrativa algorítmica, en GAMERO CASADO, E. (Dir.), Inteligencia artificial y sector público. Retos, límites y medios, Tirant Lo Blanch, Valencia.
-MARTÍNEZ GUTIÉRREZ, R. (2023), Responsabilidad patrimonial por el uso de Inteligencia Artificial, en GAMERO CASADO, E. (Dir.), Inteligencia artificial y sector público. Retos, límites y medios, Tirant Lo Blanch, Valencia.
-MARTÍNEZ MARTÍNEZ, R. (2019), Inteligencia artificial desde el diseño. Retos y estrategias para el cumplimiento normativo, RCDP, nº 58.
-MERCHÁN ARRIBAS, M. (2020), Guía de Uso de la Inteligencia Artificial en el Sector Público, Club de la Innovación, disponible en file:///C:/Users/MatildeCarl%C3%B3n/Downloads/guia-uso-de-la-ia_en-el-sector-publico%20(3).pdf
-MIR PUIGPELAT, O. (2023), Algoritmos, Inteligencia Artificial y procedimiento administrativo: principios comunes en el Derecho de la Unión Europea, en GAMERO CASADO, E. (Dir.), Inteligencia artificial y sector público. Retos, límites y medios, Tirant Lo Blanch, Valencia.
-MISURACA, G. and VAN NOORDT, C. (2020), AI Watch - Artificial Intelligence in public services, EUR 30255 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-19540-5, doi:10.2760/039619, JRC120399.
-NIETO, A. (2021, El arbitrio judicial. Entrando en la mente del juez, 3ª ed., Colex, A Coruña.
-NIKOLINAKOS, N. (2023), U Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies - The AI Act, Springer.

-PONCE, J. (2018), Inteligencia artificial, Derecho administrativo y reserva de humanidad: algoritmos y procedimiento administrativo debido tecnológico, RGDA, nº 50.

(2022), Reserva de humanidad y supervisión humana de la Inteligencia Artificial, El Cronista del Estado social y democrático de Derecho, nº 100.

-PRESNO LINERA, M. A. (2023), Derechos fundamentales e Inteligencia Artificial, Tecnos, Madrid.

- QUINTAVILLA, A. y TEMPERMAN, J. (Eds.) (2003), Artificial Intelligence and Human Rights, Oxford University Press.

-ROBLES MARTÍN-LABORDA, C. (2018), Cuando el cartelista es un robot. Colusión en mercados digitales mediante algoritmos de precios, Actas de Derecho Industrial, Vol. 18.

- SORIANO ARNAIZ, A. (2021), Decisiones automatizadas: problemas y soluciones jurídicas. Más allá de la protección de datos, Revista de Derecho Público: Teoría y Método, Vol. 3.

-TORRECILLA-SALINAS, C.; TANGI, L.; ULRICH, P.; MANZONI, M.; SCHADE, S.; MARTÍNEZ-RODRÍGUEZ, E. y PIGNATELLI, F. (2023), ¿Para qué sirve la Inteligencia Artificial en el sector público? Casos de uso y perspectivas de aplicación, en GAMERO CASADO, E. (Dir.), Inteligencia artificial y sector público. Retos, límites y medios, Tirant Lo Blanch, Valencia.

-VALERO TORRIJOS, J. (2023), Las singularidades del tratamiento de datos de carácter personal en entornos de inteligencia artificial en el sector público, en GAMERO CASADO, E. (Dir.), Inteligencia artificial y sector público. Retos, límites y medios, Tirant Lo Blanch, Valencia.

-VALOR YÉBENES, J.A. (2023), Humanismo, Transhumanismo e Inteligencia Artificial.

-VELASCO RICO, C. I. (2020), Personalización, proactividad e inteligencia artificial. ¿Un nuevo paradigma para la prestación electrónica de servicios públicos?, IDP, nº 30.

-VESTRI, G. (2021), La inteligencia artificial ante el desafío de la transparencia algorítmica. Una aproximación desde la perspectiva jurídico-administrativa, en RArAP, nº 56.

-YUSTE, R. (2019), Las nuevas neurotecnologías y su impacto en ciencia, medicina y sociedad, Lecciones Cajal, Universidad de Zaragoza, disponible en https://puz.unizar.es/2183-las-nuevas-tecnologias-y-su-impacto-en-la-ciencia-medicina-y-sociedad.html

-ZLOTNIK, A. (2019), Inteligencia Artificial en las Administraciones Públicas: definiciones, evaluación de viabilidad de proyectos y áreas de aplicación, Boletic, nº 84.

DOCUMENTOS CITADOS

- AGENCIA ESPAÑOLA DE PROTECCIÓN DE DATOS (AEPD) (2016), Orientaciones y garantías en los procedimientos de anonimización de datos personales, disponible en https://datos.gob.es/es/documentacion/orientaciones-y-garantias-en-los-procedimientos-de-anonimizacion-de-datos-personales

(2018), Orientaciones para prestadores de servicios de *cloud computing*, disponible en https://www.aepd.es/documento/guia-cloud-prestadores.pdf

(2020), Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción, disponible en https://www.aepd.es/prensa-y-

comunicacion/notas-de-prensa/la-aepd-publica-una-guia-para-adaptar-al-rgpd-los-productos-y

(2021a), 10 malentendidos relacionados con la anonimización, disponible en https://www.aepd.es/documento/10-malentendidos-anonimizacion.pdf

(2021b), Requisitos para auditorías de tratamientos que incluyan Inteligencia Artificial, disponible en https://www.aepd.es/prensa-y-comunicacion/notas-de-prensa/aepd-publica-guia-requisitos-auditorias-tratamiento-ia

(2022c), Gestión de riesgos y evaluación de impacto en tratamiento de datos personales, disponible en https://www.aepd.es/documento/gestion-riesgo-y-evaluacion-impacto-en-tratamientos-datos-personales.pdf

(Sin fecha), Lista orientativa de tipos de tratamientos que no requieren una evaluación de Impacto relativa a la protección de datos según el artículo 35.5 RGPD, disponible en https://www.aepd.es/documento/listasdpia-35.5l.pdf

(Sin fecha), Lista de tipos de tratamientos de datos que requieren evaluación de impacto relativa a la protección de datos (art. 35.4), disponible en https://www.aepd.es/documento/listas-dpia-es-35-4.pdf.

(Sin fecha), Plantilla para la elaboración de la Evaluación particularmente por parte de entidades del sector público, disponible en https://www.aepd.es/derechos-y-deberes/cumple-tus-deberes/medidas-de-cumplimiento/evaluaciones-de-impacto

-AGENCIA ESPAÑOLA DE PROTECCIÓN DE DATOS, AGENCIA CATALANA DE PROTECCIÓN DE DATOS y AUTORIDAD VASCA DE PROTECCIÓN DE DATOS (Sin fecha), Directrices para la elaboración de contratos entre el responsable y el encargado del tratamiento, disponible en https://www.aepd.es/documento/guia-directrices-contratos.pdf

-CARTA DE DERECHOS DIGITALES, Documento sin fecha, disponible en https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/140721-Carta_Derechos_Digitales_RedEs.pdf

-CENTRAL DIGITAL DATA OFFICE (2019), A guide to using artificial intelligence in the public sector, https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector

-COMISIÓN EUROPEA (2018), Comunicación "Inteligencia Artificial para Europa" [COM (2018) 237 final, de 25.4.2018]

(2019), Comunicación al Parlamento Europeo, al Consejo, al Comité Económico y Social y al Comité de las Regiones "Generar confianza en la Inteligencia Artificial centrada en el ser humano" [COM (2019)168, de 8 de abril de 2019]

(2020), Libro Blanco de la UE sobre la inteligencia artificial: un enfoque europeo orientado a la excelencia y a la confianza, Comunicación de la Comisión Europea [COM(2020) 65 final, de 19 de febrero]

-CONSEJO DE LA OCDE (2019a), Recomendaciones sobre la Inteligencia Artificial, mayo de 2019, disponible en https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

(2019b), *Artificial Intelligence in Society*, 11 de junio de 2019, hps://www.oecd-ilibrary.org/sites/eedfee77-en/1/2/1/index.html?itemId=/content/publication/eedfee77-

-ESTRATEGIA NACIONAL DE INTELIGENCIA ARTIFICIAL (2020), disponible en https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/ENIA2B.pdf

-GRUPO INDEPENDIENTE DE EXPERTOS DE ALTO NIVEL SOBRE INTELIGENCIA ARTIFICIAL (HLEG-AI) (2019), Guías Éticas para una IA fiable, disponibles en https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai)

-PUBLIC BUYERS COMMUNITY (2023), Propuesta de cláusulas contractuales tipo para la contratación de inteligencia artificial por parte de organismos públicos, disponibles en la página del Observatorio de Contratación Pública: https://www.obcp.es/noticias/nueva-version-de-las-clausulas-de-contratacion-publica-de-la-ia-apoyo-al-uso-responsable
- SAMOILI, S., LÓPEZ COBO, M., GÓMEZ GUTIÉRREZ, E., DE PRATO, G., MARTÍNEZ-PLUMED, F. y DELIPETREV, B., AI WATCH. Defining Artificial Intelligence, EUR 30117 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-17045-7, doi:10.2760/382730, JRC118163, accesible en file:///C:/Users/MatildeCarl%C3%B3n/Downloads/jrc118163_ai_watch._defining_artificial_intelligence_1.pdf
-SERVICIOS DEL PARLAMENTO EUROPEO, Informe sobre el Proyecto de Reglamento de Inteligencia Artificial, junio de 2023, disponible en https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf
-TANGI, L.; VAN NOORDT, C.; COMBERTO, M.; GATTWINKEL, D. y PIGNATELLI, F., AI Watch. European Landscape on the use of Artificial Intelligence by the Public Sector, EUR 31088 EN, Publications Office on the European Union, Luxembourg, 2022, ISBN 978-92-76-53058-9, doi: 10.2760/39336, JRC129301, accesible en https://publications.jrc.ec.europa.eu/repository/handle/JRC129301
- WIERZBOWSKI, M. (Coord.) (2022), Model Rules on Impact Assesment on Algorithm Decision-making Systems Used by Public Administrations. Report of the European Law Institute, Wien Universität.
- WOLSWINKEL, J. (2022), Artificial Intelligence and Administrative Law. Comparative study, Council of Europe, disponible en https://www.coe.int/en/web/cdcj/-/new-report-on-the-interplay-between-artificial-intelligence-and-administrative-law