TEXTO PROVISIONAL

Juan Pedro Quintana Carretero. Magistrado del Tribunal Supremo

TRANSPARENCIA Y DERECHO DE ACCESO A LA INFORMACIÓN PÚBLICA. ESPECIAL REFERENCIA A LA ACTIVIDAD AUTOMATIZADA DE LA ADMINISTRACIÓN PUBLICA.

I. INTRODUCCION

Antes de abordar el examen de la transparencia algorítmica y el empleo de la inteligencia artificial en las Administraciones públicas, incluida la Administración de Justicia, resulta necesario hacer algunas consideraciones generales sobre la inteligencia artificial con el objeto de comprender su significado y alcance.

Cabe preguntarse por el concepto de inteligencia artificial

Se trata de un término compuesto por dos palabras que no reflejan fielmente su significado:

En primer lugar, inteligencia, según una acepción de la RAE, es la “capacidad de resolver problemas”. Dicho de otra forma, es la capacidad de encontrar soluciones para nuevos problemas o la capacidad de resolver problemas complejos y adaptarse a nuevas situaciones. Desde esta perspectiva cabe afirmar que la IA es inteligente.

No obstante, en otra acepción de inteligencia de la RAE es la “capacidad de entender y comprender”, y desde esta perspectiva no puede afirmarse que la IA sea, verdaderamente, inteligente.

La palabra inteligencia proviene del latín *intelligentia o intellēctus* que significa 'comprender' o 'percibir'. Existen numerosas acepciones del término inteligencia y muchas de ellas están unidas o conectadas con nociones básicas como lógica, comprensión, autoconciencia, aprendizaje, conocimiento emocional, razonamiento, planificación, creatividad, pensamiento crítico y resolución de problemas. En términos más generales podemos decir que es la capacidad para percibir o inferir información y retenerla como conocimiento para aplicarla para la resolución de problemas.

Pues bien, la IA no es inteligencia porque, a diferencia de la inteligencia física, (i) no se aplica a todas las actividades del ser humano; (ii) no tiene componentes intuitivos y emocionales vinculados a experiencias personales o contextos (sociales, económicos, históricos o culturales) -solo tiene datos-, que inciden en la toma decisiones humanas, y no opera causalmente, sino a través de complejos procesos de inferencia de resultados a partir de datos -inducción y deducción-, y (iii) su método de razonamiento difiere del propio del pensamiento humano pues la IA se limita a hacer inferencias inductivas -consistentes en determinar valores- y deductivas -que desarrolla las consecuencia necesarias de una hipótesis pura-, no abductivas reservadas a los seres humanos.

En definitiva, aunque la IA generativa pretende emular el comportamiento inteligente humano, no responde del todo a este concepto pues se limita a gestionar y analizar un volumen de datos enorme bajo un método estadístico, buscando y localizando donde se emplean los términos o palabras que escribimos en nuestras preguntas, para ofrecer respuestas, siguiendo una metodología inductiva o deductiva, no abductiva.

Las inferencias abductivas permiten realizar conjeturas vinculadas a una comprensión global del mundo y basadas en las experiencias vividas, es decir, consisten en llegar a una conclusión probable a partir de lo que sabemos. Este proceso decisorio implica la existencia de empatía pues requiere considerar una aspecto o punto de vista singular implicado en el razonamiento -sin empatía el razonamiento abductivo es imposible-.

Puede concluirse que los sistemas de IA manifiestan un comportamiento inteligente solo en la medida que son capaces de analizar el entorno y llevar a cabo una acción con cierto grado de autonomía, con el fin de alcanzar objetivos específicos, y pueden consistir en un simple programa informático, como motores de búsqueda o sistemas de reconocimiento facial o de voz, o en complejos programas incorporados a dispositivos de hardware como robots o automóviles autónomos.

En segundo lugar, artificial es todo aquello que no es natural y que es hecho con intención y propósito. La IA tampoco es únicamente artificial, porque, aunque es una creación humana, se encuentra controlada por el ser humano, pues funciona y es aplicada y supervisada por humanos, no funciona autónomamente en sentido estricto, no opera como un ente autónomo dotado de razón.

En definitiva, aunque la IA puede aprender de su propia experiencia, como ocurre con los seres humanos, no puede pensar y actuar por si misma, pues su funcionamiento se basa en un procedimiento de tratamiento estadístico de una ingente cantidad de datos, capaz de ser producir previsiones con alto grado de acierto y con cierta capacidad de aprendizaje. Los sistemas de IA son sistemas de información sobre los que se ejecutan procesos informáticos (algoritmos) de aprendizaje y decisión.

Por ello, podemos afirmar que la IA no es creativa, no tiene imaginación y no tiene capacidad para tomar decisiones por si misma fuera de los límites de su programación, fuera de lo que ha sido programada para hacer.

Concluimos, por tanto, que la ausencia de capacidad de la IA para ser creativa o tener libre albedrio es una de las principales diferencias entre ella y los seres humanos

La IA es una rama de la informática que tiene como objetivo crear sistemas capaces de realizar tareas que tradicionalmente requieren inteligencia humana que incluyen reconocer patrones, tomar decisiones, resolver problemas y aprender de la experiencia.

Verdaderamente, la IA no responde a una sola técnica, sino que es una familia de técnicas que son utilizadas para simular o recrear la inteligencia humana a través de las ciencias computacionales con la finalidad de que los sistemas dotados de una gran velocidad de procesamiento y de clasificación de datos, proporcionen respuestas o propuestas rápidas a las preguntas o problemas que se formulen -es mucho más que un mero fenómeno técnico-.

La inteligencia artificial no es una tecnología única, sino un conjunto de herramientas y métodos que buscan simular la capacidad cognitiva humana, permitiendo que las máquinas que interpreten, procesen y actúen de acuerdo con información completa. Comprende campos tan diversos como el procesamiento del lenguaje natural o el aprendizaje automático y se encuentra en constante evolución.

Dejando a un lado la perspectiva meramente tecnológica, desde un punto de vista jurídico, por inteligencia artificial, podemos entender, siguiendo la definición de Sistema de Inteligencia Artificial del Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, por el que se establecen normas armonizadas en materia de inteligencia artificial (artículo 3.1) como aquel que opera con elementos de autonomía y que, basándose en datos y entradas obtenidos de humanos o máquinas, infiere como alcanzar unos objetivos propuestos, usando para ello técnicas basadas en el aprendizaje automático o en lógica y conocimiento, y genera como salida contenidos, predicciones, recomendaciones o decisiones que influyen en el entorno con el que el sistema interactúa.

En definitiva, es un sistema o programa informático que tiene objetivos definidos y resultados predeterminados, consistentes en contenidos, predicciones, recomendaciones, decisiones o aprendizaje, con la capacidad de influir en el entorno y cierto grado de autonomía para emprender acciones.

A su vez, los sistemas de IA pueden usarse con IA conexionista, no simbólica o estadística (sistemas de aprendizaje automático o profundo que realizan inducciones mediante métodos estadísticos aplicados a datos -*machine learning*-) o con IA simbólica o no estadística (IA basada en reglas predeterminadas). Además, pueden emplearse como mero apoyo a la decisión administrativa, interviniendo como un *instrumento*, o para tomar la decisión administrativa de forma automatizada, sin intervención humana, operando como un *agente*.

Así, debemos distinguir entre inteligencia artificial simbólica y aprendizaje automático o *machine learning*.

La inteligencia artificial simbólica es una rama de la inteligencia artificial que se basa en reglas lógicas predefinidas y en conocimiento explícito introducido por humanos. A diferencia de la rama del aprendizaje automático, no aprende de los datos, sino que opera mediante la manipulación de símbolos y la aplicación de reglas formales para representar el conocimiento y razonar sobre él.

Este tipo de IA fue uno de los primeros enfoques en desarrollarse y tiene sus raíces en la lógica matemática y la filosofía. Utiliza estructuras como sistemas de reglas, ontologías, árboles de decisión y motores de inferencia que permiten a las máquinas realizar deducciones, responder preguntas o resolver problemas en base a una base de conocimientos claramente estructurada. Por ello, no puede adaptarse a nuevos datos o factores inesperados.

Sin embargo, el aprendizaje automático es una rama de la inteligencia artificial en la que los sistemas aprenden de los datos para hacer predicciones y suele entrenarse con modelos de datos históricos.

A su vez, existen varios tipos de algoritmos que emplean aprendizaje automático: regresión lineal, árboles de decisión, redes neuronales simples, aprendizaje no supervisado (algoritmos comunes: Clustering o agrupamiento, reducción de la dimensionabilidad -PCA-) y aprendizaje por refuerzo (Q-Learning).

Entre ellos, el *Deep learning* o aprendizaje profundo es una subdisciplina del *machine learning* que utiliza redes neuronales profundas, es decir, redes con muchas capas, y se utiliza cuando hay grandes volúmenes de datos y relaciones complejas entre las diferentes variables. Su principal inconveniente es que, aunque es muy preciso en sus resultados, no es fácil de entender cómo cada característica o cada dato afecta la predicción que se ha hecho ya que el proceso es relativamente opaco.

Aunque las investigaciones sobre inteligencia artificial comenzaron a mediados del siglo pasado, realmente es en 2017 cuando surge la IA generativa *(machine learning)* que utiliza las redes de neuronales profundas con capacidad de aprendizaje de la experiencia (algoritmos basados en múltiples datos de redes neuronales) y son capaces de crear o generar imágenes, videos, textos, traducir idiomas.

Las investigaciones sobre IA generativa comenzaron en la Conferencia de Dartmouth en 1956 (Hannover, New Hampshire, Estados Unidos), que reunió a un grupo de científicos para discutir sobre esta disciplina y el uso del ordenador para diseñar conductas inteligentes. De este modo se inició la investigación sobre maquinas generativas que pensaban y mejoraban por si mismas y su propósito era introducir capacidad cognitiva en una máquina.

Desde un punto de vista técnico, el desarrollo de la IA generativa ha sido posible gracias a la introducción de nuevas arquitecturas de redes neuronales y aprendizaje profundo que permiten avanzar en la creación de imágenes y la predicción en la generación de lenguaje junto con un entrenamiento eficiente a partir de grandes conjuntos de datos.

En general, la IA generativa utiliza cualquier dato digitalizado (palabras, imágenes, datos biométricos, voz, ubicaciones, etc.) y los gestiona de diferentes maneras. Su característica más relevante es que es capaz de aprender y generar una IA mejor, es decir, no solo agrega y conecta datos sino que aprende de su propio funcionamiento, con o sin supervisión humana.

Por ello, el funcionamiento de la IA generativa se basa en la incorporación al sistema de datos o información (corpus), hasta el punto de que resulta

fundamental para su buen funcionamiento la adecuada selección de tales datos o informaciones -anonimizados, veraces y actualizados-, dado que son los elementos sobre los que se construye la información y de los que deriva el conocimiento.

Por tanto, los datos son el elemento central de la IA y sin datos de calidad no se pueden desarrollar sistemas de IA que funcionen correctamente para lograr un objetivo determinado, pues alimentan los algoritmos de aprendizaje y permiten crear modelos más complejos y con una mayor precisión. Ello explica que se haya afirmado que "los datos son el petróleo del siglo XXI", destacando su importancia y el valor en la economía y la sociedad moderna.

Precisamente, la relevancia de los datos hace que la manipulabilidad de la IA sea un riesgo real y dependerá de los corpus o textos que alimenten el sistema. Por ello, en Europa se ha creado una autoridad europea de supervisión, al margen de las que se creen en los EEMM, para supervisar esos corpus y controlar el funcionamiento histórico del algoritmo, comprobando si es fiel al objetivo para el que fue creado y si cumple las normas llamado a aplicar, entre otros aspectos. Se trata de un control complejo, cuya eficacia exige la conservación de toda la trazabilidad del proceso de elaboración del sistema de IA y la identificación de los responsables en cada paso o etapa del proceso

Aunque existen herramientas de IA generativa específicas y propósito general, las modalidades de IA generativa cuyo uso se encuentra más extendido socialmente son las de propósito general, dada su versatilidad, al poderse emplear en cualquier dominio (jurídico, medio ambiente, seguros, sanidad, etc.), como ocurre con Gemini de Google, Chatgpt de Open AI o Claude de Antropic. Hoy estas herramientas están presentes en todos los ámbitos de la sociedad y la economía, y las previsiones indican que pocos aspectos quedarán fuera de su influencia en los próximos años.

Al respecto se ha dicho que la IA generativa, considerada como tecnología de propósito general, está impulsando la cuarta revolución industrial y que su impacto en la sociedad puede compararse con el de la electricidad a principios del siglo pasado.

Por ejemplo, se emplea en la medicina para diagnósticos mediante el análisis inteligente de imágenes, análisis de señales de electrocardiogramas, etc, en las empresas para analizar el comportamiento de los clientes y hacer predicciones, en el uso diario de Spotify o Netflix.

Y a todo ello se une que junto a la IA generativa, ya se van implementando progresivamente modelos de IA más potentes como la computación neuromórfica, que intenta imitar el proceso mental humano, o la IA neurosimbólica, que combina capacidades de aprendizaje estadístico y basado en datos de las redes neuronales con el razonamiento simbólico

El siguiente paso evolutivo será la IA física (robotización) para la realización de las tareas propias del ser humano con IA que aún no está tan desarrollada para justificar la generalización de su uso en la vida cotidiana, pero sobre la que se dice que en menos de una década se incorporará con normalidad a nuestras vidas.

Más lejos aún será posible llegar con el desarrollo de la convergencia entre la neurotecnología y la inteligencia artificial que está produciendo una mutua aceleración de ambas tecnologías, con la que se pretende alcanzar un mejor conocimiento de la naturaleza de la mente humana y enormes utilidades en el ámbito de la salud. Los avances de la neurotecnología permiten descodificar los procesos electroquímicos que configuran la actividad cerebral, que pueden ser interferidos, induciendo pensamientos, expresiones, emociones o acciones. De modo que esta convergencia tecnológica conlleva que los neurodispositivos se desarrollen en interacción con la inteligencia artificial, aprovechando la capacidad de esta última de gestionar una enorme cantidad de datos.

Dado el impacto de la IA en nuestras vidas, cabe preguntarse ¿cómo afectará la evolución de la IA a la percepción, la cognición y la interacción humana? ¿Y cuál será el impacto de la IA en nuestra cultura, nuestro concepto de humanidad y, en definitiva, nuestra historia?

En fin, la IA no es un fenómeno meramente técnico y su aparición no es provisional. Se trata de una revolución tecnológica con efectos transversales que avanzará paulatinamente hacia la consolidación de la IA como herramienta de uso generalizado que la inmensa mayoría de los ámbitos de la actividad humana. Y, todo ello a una enorme velocidad, lo que supone un verdadero reto de adaptación de los sistemas políticos y nuestra sociedad a este nuevo escenario.

Sin duda, la IA ha llegado para quedarse y penetrar en muy diferentes áreas de conocimiento con los consiguientes efectos, también, sobre la prestación de variados servicios públicos y, como no, en la Administración de Justicia. En estos escenarios resulta evidente que su uso afecta a pilares esenciales del Estado de derecho: el principio de legalidad, la tutela judicial efectiva, el derecho de defensa y la igualdad ante la ley.

Por ello, no faltan quienes afirman que la IA puede poner en peligro el Estado de Derecho y los sistemas políticos democráticos.

El Center for AI Safety, organización estadounidense sin ánimo de lucro, con sede principal en San Francisco, dedicada a estudiar y reducir los riesgos graves o de gran escala asociados al desarrollo de la inteligencia artificial, promovió en 2023 la llamada “Statement on AI Risk”, una declaración breve en la que se afirmaba que: "Mitigar el riesgo de extinción derivado de la IA debería ser una prioridad mundial, junto con otros riesgos a escala social como las pandemias y la guerra nuclear". La declaración fue firmada por científicos, empresarios y figuras relevantes del sector tecnológico. La declaración fue relevante porque no la suscribieron solo críticos externos de la IA, sino también directivos y científicos vinculados directamente a OpenAI, Google DeepMind, Anthropic, Microsoft y universidades de máximo prestigio.

Ciertamente, no son desdeñables los riesgos que entraña su uso, especialmente por su mal uso, ante las posibilidades de manipulación de la opinión pública mediante la generación y difusión en noticias falsas o información manipulada incluso mediante campañas destinadas a mermar a erosionar la confianza de los ciudadanos de las instituciones, o a través del uso de datos personales para combatir la discrepancia con las clases políticas gobernantes.

Sin embargo, no son desdeñables tampoco las ventajas que puede suponer para el fortalecimiento de los sistemas democráticos pues su potencialidad en el análisis de ingentes cantidades de datos posibilita un eficaz examen de la información pública y facilita su acceso a los ciudadanos incrementando la información a su disposición y, en definitiva, favoreciendo la transparencia.

Sin duda en estos tiempos, donde es creciente la desafección de los ciudadanos de los de las instituciones públicas propias de las democracias occidentales, un adecuado sistema de rendición de cuentas y transparencia resulta vital para fortalecer esa confianza.

En este escenario las herramientas de IA, además de mejorar la eficiencia de los procesos burocráticos y reducir los tiempos de respuesta de las Administraciones Públicas, a través del diagnóstico y análisis de los datos necesarios en la adopción de decisiones, permitiría paliar la deslegitimación progresiva de los sistemas políticos democráticos, por ejemplo, facilitando la rendición de cuentas y la transparencia y fomentando el uso de procesos participativos mediante la creación de plataformas digitales que permitieran la participación de los ciudadanos en la toma de decisiones.

Cabe también sostener que puede favorecer la igualdad social (evitando sesgos en decisiones administrativas), procesos democráticos más participativos y transparentes y un mejor funcionamiento de las instituciones públicas.

Es evidente que la IA entraña importantes riesgos que debemos anticipar y gestionar, pero es innegable que tiene importantes ventajas para la sociedad y ya se emplea con normalidad en variadísimos ámbitos públicos y privados.

El Papa León XIV ha mostrado su preocupación por las consecuencias sociales de uso de la IA en la encíclica “Magnifica Humanitas”, reclamando responsabilidad identificable, transparencia, posibilidad de recurrir las decisiones y reparación de daños, mediante su sometimiento a un código ético que garantice un orden social justo en la era digital, un marco jurídico adecuado y mecanismos de protección eficaces.

De ahí la importancia de la gobernanza del riesgo y las nuevas regulaciones sobre el particular, cuya interpretación y aplicación requiere a los juristas conocimientos sobre esta tecnología, necesarios para la correcta aplicación de esas normas, comprendiendo su sentido y alcance. Pero lo cierto es que el desarrollo tecnológico de la IA parece imparable y es posible que sea más rápido

que nuestra capacidad para prevenir y evitar sus riesgos, pues la experiencia demuestra que la regulación suele ir por detrás de la sociedad y sus avances tecnológicos.

La Administración de Justicia no pueden ser ajena al desarrollo de la “inteligencia generativa”, ni a su eventual aplicación en los distintos órdenes jurisdiccionales, si bien deben ponderarse los problemas reales de su implementación y los riesgos de su funcionamiento, asegurando su fiabilidad. Al respecto, es importante no olvidar la perspectiva ética, soslayar los eventuales sesgos discriminatorios, respetar la privacidad y la confidencialidad y asegurar la fiabilidad de su empleo. En definitiva, su implantación debe estar guiada por la responsabilidad, la prudencia y la proporcionalidad.

La implementación de tecnologías algorítmicas y la IA generativa obliga a toda la judicatura a repensar su rol en una era donde las decisiones automatizadas, los sistemas predictivos y la prueba generada por máquinas se abren paso en los tribunales. Mas que nunca, es necesaria una judicatura formada y, además hoy vigilante y con autonomía epistemológica frente a la opacidad algorítmica.

Entre las cuestiones que hemos de abordar en el control judicial sobre los poderes públicos que emplean estos sistemas de IA se encuentra la vigencia del principio de transparencia y el acceso a la información pública, entendida en los Estados democráticos como un bien público que debe estar al alcance de los ciudadanos y sirve al control institucional sobre los representantes de la sociedad. De manera que la transparencia de las Administraciones públicas sobre la gestión de los asuntos públicos es una característica sustancial a los Estados democráticos.

Por tanto, el derecho de acceso a la información pública es consustancial a un régimen democrático, contribuye a la formación de una opinión pública, libre e informada y favorece la participación de los ciudadanos en el control del poder.

En este examen debemos atender al marco regulatorio vigente, más allá del derecho interno, en particular, el Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (RIA), uno de cuyos aspectos relevantes es el equilibrio que muestra entre el fomento de la innovación en el ámbito de la IA y el establecimiento de cautelas para proteger a las personas frente al mal uso de la IA, equilibrio de difícil consecución, como pone de manifiesto la extraordinaria complejidad y escasa claridad del RIA que cuenta con 113 artículos y XIII anexos.

En efecto, la IA es una herramienta poderosa con un enorme potencial, pero junto a sus beneficios han aparecido riesgos importantes relacionados con la privacidad, la seguridad y el respeto de los derechos fundamentales, lo que ha

impulsado la necesidad de un régimen regulatorio que garanticen el uso seguro de estas tecnologías.

El examen de esta cuestión exige comentar y extraer consecuencias de la STS num. 1119/2025, de 11 de septiembre de 2025 (rec. 7878/2024), denominada caso BOSCO, objeto de atención por la doctrina científica en decenas de comentarios publicados, mayoritariamente favorables, que han generado un intenso e interesante debate sobre la cuestión de la transparencia algorítmica, no solo a nivel nacional sino también internacional.

Esta sentencia gira en torno a tres ideas troncales:

1.- Aborda una interpretación dinámica de un derecho constitucional, sustentado en el artículo 105 CE y el artículo 42 de la Carta de Derechos Fundamentales de la Unión Europea, a cuya luz interpreta y aplica el artículo 14 LTAIBG, configurando el derecho de acceso a la información pública como un derecho constitucional subjetivo ejercitable frente a las Administraciones Públicas, que tiene un carácter instrumental vinculado a derechos fundamentales, y deriva de exigencias de democracia y transparencia, inseparablemente unido al Estado Democrático y de Derecho.

2.- Resalta la especial relevancia de los riesgos que para los ciudadanos entraña el uso por las Administraciones Públicas de sistemas informáticos en la toma de decisiones administrativas automatizadas, que deben conllevar exigencias de transparencia de los procesos informáticos empleados. Lo cual se conecta con el principio de buena administración.

3.- Acuña el principio de «transparencia algorítmica» como manifestación específica del derecho constitucional de acceso cuando su objeto son algoritmos o códigos fuente de aplicaciones administrativas.

La sentencia sustenta el acceso al código fuente del algoritmo, prescindiendo de la aplicación directa de normativa específica sobre digitalización o inteligencia artificial pues no resultaba aplicable para resolver sobre la pretensión de acceso a información pública ejercitada por la fundación Bosco, situada cronológicamente en año 2018, y se sustenta en el derecho constitucional de acceso a la información pública y la LTAIPBG. Frente a concepciones reduccionistas que limitan la transparencia algorítmica a explicaciones funcionales, la sentencia admite expresamente que «puede requerir el acceso a su código fuente» cuando resulte necesario para verificar la conformidad del algoritmo con las previsiones normativas.

Volveremos sobre esta sentencia más adelante.

Concluimos esta introducción, señalando que la inteligencia artificial apenas cuenta con regulación en nuestro ordenamiento jurídico, con la excepción de

alguna norma estatal y alguna otra de las comunidades autónomas, residiendo su regulación más importante en el reglamento europeo de inteligencia artificial.

No obstante, hemos de hacer alguna referencia a sus fundamentos constitucionales en la medida en que pueda tener incidencia en los derechos fundamentales y en la regulación de los principios rectores de la administración pública.

Así, en primer lugar, el artículo 18.4 de la Constitución establece que la ley limitará el uso de la informática para garantizar el honor y la intimidad personal y familiar de los ciudadanos y el pleno ejercicio de sus derechos: Este precepto justifica el establecimiento de límites en el uso de la inteligencia artificial mediante ley.

En segundo lugar, el artículo 105.1 de la Constitución dispone que la ley regulará el procedimiento a través del cual debe producirse los actos administrativos, así como el acceso de los ciudadanos a la información pública, lo que nos conduce a una reserva legal para la incorporación de los sistemas de inteligencia artificial en el procedimiento administrativo y a la vigencia del principio de transparencia en los sistemas de inteligencia artificial y su uso por las Administraciones Públicas.

Debe ponerse de manifiesto que las leyes que regulan el procedimiento administrativo común de las Administraciones Públicas y el régimen jurídico del sector público, leyes 39/2015 y 40/2015, ambas de 1 de octubre, así como él reglamento de actuación y funcionamiento del sector público por medios electrónicos, el Real decreto 203/2021, de 30 de marzo, no hacen referencia alguna a la inteligencia artificial.

Por lo que respecta a la normativa estatal, el uso de la inteligencia artificial por la Administración pública solo se menciona en el artículo 23 de la ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación.

II. ALGUNOS CONCEPTOS TÉCNICO-JURIDICOS

Antes de proseguir, para facilitar el examen de esta materia, conviene aproximarnos a estos conceptos técnico-jurídicos, no siempre de fácil comprensión, enunciando sus rasgos fundamentales. Veamos:

1.- Los **algoritmos** incorporados a las aplicaciones o programas informáticos, según el Diccionario de la Real Academia Española, son un «conjunto ordenado y finito de operaciones que permite hallar la solución de un problema», es decir, un grupo finito de operaciones organizadas de manera lógica y ordenada que permiten alcanzar la solución a un problema determinado o realizar una tarea específica mediante un ordenador o computadora. De modo que funcionan

mediante una cadena de instrucciones preestablecidas que determinan el seguimiento de unos determinados pasos programados hasta alcanzar el resultado pretendido.

2.- El **código fuente del algoritmo** expresa esas operaciones descritas en el lenguaje de la programación, es decir, es la traducción concreta de un algoritmo a un lenguaje de programación que puede ser entendido y ejecutado por un ordenador o computadora. De modo que constituye la representación escrita del algoritmo en un lenguaje de programación, sin perjuicio de que las operaciones así descritas resulten traducibles a lenguaje humano.

Dicho de otra manera, empleando las palabras del Consejo de Transparencia, 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.

3.- Con motivo del empleo por las Administraciones públicas de sistemas de IA y decisiones administrativas automatizadas, surge el llamado **principio de "transparencia algorítmica",** que impone a las Administraciones públicas obligaciones de información pública para facilitar el acceso de los ciudadanos, en mayor o menor medida, a las características fundamentales de los algoritmos empleados en la toma de decisiones o su código fuente, como una manifestación del principio de transparencia, consagrado constitucionalmente (artículo 105.b) de la CE).

4.- En íntima conexión con este principio aparece un concepto de mayor amplitud: **"democracia digital o electrónica".** Nace como consecuencia del uso de las tecnologías digitales por los gobiernos y los ciudadanos, y su desarrollo pretende fortalecer las prácticas democráticas tradicionales. La democracia digital no solo es una extensión tecnológica de la democracia representativa, sino que también es el fruto de una auténtica transformación estructural en el funcionamiento democrático de los Poderes públicos, caracterizada por la vigencia de los principios de transparencia, participación y rendición de cuentas en un entorno digital, donde el acceso a la información púbica y la transparencia algorítmica ocupan un papel esencial para garantizarla.

En este nuevo contexto digital democrático se impone a los Poderes públicos la obligación, entre otras, de explicar de forma comprensible el funcionamiento de los algoritmos que se emplean en la toma de decisiones que afectan a los ciudadanos para permitirles conocer, fiscalizar y participar en la gestión pública.

5.- Las **actuaciones administrativas automatizadas** deben ser entendidas como «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» (artículo 41.1 de la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público).

III. LA TRANSPARENCIA Y EL ACCESO A LA INFORMACIÓN PÚBLICA. MARCO NORMATIVO Y JURISPRUDENCIA

III.1.- El marco normativo del derecho de acceso a la información pública:

1.- La **Carta de los Derechos Fundamentales de la Unión Europea** reconoce el derecho de acceso a los documentos de las instituciones de la Unión Europea en los siguientes términos:

«Artículo 42. Derecho de acceso a los documentos.

Todo ciudadano de la Unión y toda persona física o jurídica que resida o tenga su domicilio social en un Estado miembro tiene derecho a acceder a los documentos de las instituciones, órganos y organismos de la Unión, cualquiera que sea su soporte».

Ley Orgánica 1/2008, de 30 de julio, por la que se autoriza la ratificación por España del Tratado de Lisboa, por el que se modifican el Tratado de la Unión Europea y el Tratado Constitutivo de la Comunidad Europea, firmado en la capital portuguesa el 13 de diciembre de 2007, establece lo siguiente:

« Artículo 2. Carta de los Derechos Fundamentales de la Unión Europea .

A tenor de lo dispuesto en el párrafo segundo del artículo 10 de la Constitución española y en el apartado 8 del artículo 1 del Tratado de Lisboa , las normas relativas a los derechos fundamentales y a las libertades que la Constitución reconoce se interpretarán también de conformidad con lo dispuesto en la Carta de los Derechos Fundamentales publicada en el «Diario Oficial de la Unión Europea» de 14 de diciembre de 2007, cuyo texto íntegro se reproduce a continuación: [...]».

2.- El artículo 105.b) de la **Constitución Española** dispone:

«La ley regulará:

(...)

b) El acceso de los ciudadanos a los archivos y registros administrativos, salvo en lo que afecte a la seguridad y defensa del Estado, la averiguación de los delitos y la intimidad de las personas.».

3.- La Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno (LTAIBG), por lo que ahora nos interesa, reconoce y garantiza el acceso a la información, partiendo de la previsión contenida en el artículo 105.b) de nuestro texto constitucional.

El actual desarrollo legal del derecho de acceso a la información pública, tal y como lo prevé el artículo 105.b) de la CE, se contiene en los artículos 12 a 24 de

la Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno (LTAIBG).

III.2.- Rasgos relevantes del derecho constitucional de acceso a la información pública y su régimen legal a la luz de nuestra jurisprudencia.

El régimen normativo que establece la Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno (LTAIBG), constituye la normativa básica transversal que regula esta materia, al tiempo que complementa al resto de las normas, pero queda desplazada, actuando en este caso como supletoria, cuando otra norma legal haya dispuesto un régimen jurídico propio y específico de acceso a la información, de conformidad con lo establecido por la disposición adicional primera, apartado segundo, de la LTAIBG (vid. STS de 25 de enero de 2021 (rec. 6387/2019), FJ 4º.5).

Abordaremos a continuación, los fundamentos de la configuración constitucional del derecho de acceso a la información pública.

El derecho de acceso a la información pública se reconoce, primariamente, en el artículo 105.b) de la Constitución Española que no solo incorpora un principio objetivo rector de la actuación de las administraciones públicas, derivado de exigencias de democracia y transparencia, sino también un derecho subjetivo de las personas, ejercitable frente a las administraciones, con sujetos, objeto y límites definidos en el propio precepto constitucional (vid. STC 164/2021, de 4 de octubre, FJ 3º).

Derecho subjetivo que, aunque no tiene la consideración de derecho fundamental en atención a su caracterización y ubicación sistemática en la Constitución [vid. SSTS de 7 de febrero de 2023 (rec. 8005/2021), FJ 5º; de 21 de abril de 2023 (rec. 350/2022), FJ 3º; y de 29 de mayo de 2023 (rec. 373/2022), FJ 3º], sí se configura como un derecho constitucional, con contenido propio y efectivo que ni el legislador, ni el aplicador de la norma pueden desconocer [vid. STC 18/1981, de 8 de junio, FJ 5º, y STS de 6 de junio de 2005 (rec. 68/2002), FJ 6º], cuyo ejercicio no cabe diferir o mediatizar por remisión al ejercicio de acciones procesales (vid. STC 164/2021, de 4 de octubre, FJ 3º), y que se encuentra estrechamente vinculado con la plena efectividad de otros principios y derechos constitucionales.

En efecto, el reconocimiento constitucional de este derecho refleja una concepción de la información que obra en manos del poder público acorde con los principios inherentes al Estado democrático, en la medida que el acceso a los archivos y registros públicos implica una potestad de participación del ciudadano y facilita el ejercicio de la crítica del poder, y acorde al Estado de Derecho, en cuanto dicho acceso constituye un procedimiento indirecto para fiscalizar la sumisión de la Administración a la ley y permitir con más eficacia el control de su actuación por la jurisdicción contencioso-administrativa [vid. SSTS

de 30 de marzo de 1999 (rec. 6563/1994) y de 16 de diciembre de 2011 (rec. 4607/2009)].

Paralelamente, la jurisprudencia constitucional lo considera, junto con otros preceptos de la Carta Magna (artículos 9.2, 23.1, 27 apartados 5 y 7, 48, 125 y 129 de la CE), una de las diversas manifestaciones que contempla el texto constitucional del fenómeno participativo de los ciudadanos en las democracias actuales y al que ha sido especialmente sensible nuestro constituyente (vid. SSTC 119/1995, de 17 de julio, FJ 4º, y 175/2021, de 25 de octubre, FJ 4º).

Verdaderamente, el derecho de acceso a la información pública trasciende a su posición ordinamental y su condición de principio objetivo rector de la actuación de las Administraciones públicas, para constituir un derecho constitucional ejercitable, como derecho subjetivo, frente a las administraciones, derivado de exigencias de democracia y transparencia, e inseparablemente unido al Estado democrático y de Derecho que enuncia el artículo 1 de nuestra Constitución.

Además, se trata también de un derecho constitucional subjetivo que presenta una íntima conexión con derechos fundamentales y libertades públicas, en la medida que su ejercicio puede condicionar la plena efectividad de estos, como el derecho de participación política (artículo 23 de la CE), el derecho a la libertad de información (artículo 20 de la CE) y el derecho a la tutela judicial efectiva (artículo 24 de la CE). Esa estrecha vinculación se advierte, igualmente, con el principio de legalidad, materializado en el sometimiento de las Administraciones públicas a la Ley y al Derecho, y su salvaguarda mediante el control que los Tribunales ejercen sobre sus actuaciones, por cuanto favorece su eficaz fiscalización por la jurisdicción contencioso-administrativa.

Y, en el ámbito del Derecho internacional, que opera como pauta interpretativa conforme al artículo 10.2 de la CE, es destacable tanto el reconocimiento expreso del derecho de acceso a la información pública como derecho fundamental en sí mismo, cual sucede en el artículo 42 de la Carta de Derechos Fundamentales de la Unión Europea, donde se dispone que: «Todo ciudadano de la Unión y toda persona física o jurídica que resida o tenga su domicilio social en un Estado miembro tiene derecho a acceder a los documentos de las instituciones, órganos y organismos de la Unión, cualquiera que sea su soporte», como su vinculación y entendimiento instrumental del derecho a la libertad de expresión y a la información, como ocurre con el artículo 19.2 del Pacto Internacional de Derechos Civiles y Políticos, hecho en Nueva York el 19 de diciembre de 1966, según la Observación General CCPR/C/GC/34 del Comité de Derechos Humanos de las Naciones Unidas, pues aquel precepto que reconoce el derecho a la libertad de expresión «enuncia un derecho de acceso a la información en poder de los organismos públicos» (vid. parágrafo 18), y con el artículo 10 del Convenio Europeo de Derechos Humanos que reconoce el derecho a la libertad de expresión, conforme a la jurisprudencia del Tribunal

Europeo de Derechos Humanos que lo interpreta, a la que haremos referencia más adelante.

Sentado lo anterior sobre la configuración como derecho constitucional del derecho de acceso a la información pública, expondremos a continuación algunos rasgos relevantes de su régimen legal a la luz de nuestra jurisprudencia.

El artículo 12 de la LTAIBG reconoce el derecho de todas las personas a acceder a la información pública, entendiéndose ésta, de acuerdo con el artículo 13, como «los contenidos o documentos, cualquiera que sea su formato o soporte, que obren en poder de alguno de los sujetos incluidos en el ámbito de aplicación de este título y que hayan sido elaborados o adquiridos en el ejercicio de sus funciones». Hemos enfatizado que la delimitación subjetiva del derecho se efectúa por la Ley en términos muy amplios, reconociéndose a "todas las personas" sin mayores distinciones (vid. STS de 25 de marzo de 2021 (rec. 2578/2020), FJ 3º.4), sin necesidad de motivar la solicitud (vid. STS de 12 de noviembre de 2020 (rec. 5239/2019), FJ 4º.7, y STC 110/2018, de 4 de octubre, FJ 5º) y sin que, en todo caso, quepa excluir las solicitudes de acceso por razón del interés privado que las motiven (vid. STS de 2 de junio de 2022 (rec. 4116/2020), FJ 2º.1).

No cabe duda de que las aplicaciones o programas informáticos software- se encuentran bajo el ámbito material de la aplicación de la LTAIBG pues constituyen información pública a tal efecto, resultando irrelevante cuáles sean sus características técnicas (formato) o el material en el que se registre (soporte), cuestión esta que no resulta controvertida.

Ello no significa que se trate de un derecho ilimitado o absoluto, pero solamente puede ser limitado por los motivos predeterminados en la ley que se encuentran en los artículos 14 y 15 de la LTAIBG (vid. STC 164/2021, de 4 de octubre, FJ 3º).

El primer precepto mencionado detalla un listado de límites del derecho de acceso, que tienen por objeto la protección de los intereses que enumera el artículo y que son los siguientes: a) la seguridad nacional, b) la defensa, c) las relaciones exteriores, d) la seguridad pública, e) la prevención, investigación y sanción de los ilícitos penales, administrativos o disciplinarios, f) la igualdad de las partes en los procesos judiciales y la tutela judicial efectiva, g) las funciones administrativas, de vigilancia, inspección y control, h) los intereses económicos y comerciales, i) la política económica y monetaria, j) el secreto profesional y la propiedad intelectual e industrial, k) la garantía de la confidencialidad o el secreto requerido en procesos de toma de decisión y l) la protección del medio ambiente.

Precisa el apartado 2º del artículo 14 de la LTAIBG que la aplicación de dichos límites, cuando proceda, habrá de ser justificada y proporcionada a su objetivo y a la finalidad de protección, atendiendo a las circunstancias del caso concreto y,

especialmente, a la concurrencia de un interés público o privado superior que justifique el acceso.

Los límites contemplados en este precepto no constituyen causas de exclusión (vid. STS de 16 de diciembre de 2019 (rec. 316/2018), FJ 4º.C) ni la apreciación de su concurrencia es una potestad discrecional de la Administración (vid. STS de 29 de mayo de 2023 (rec. 373/2022), FJ 4º), ni cabe su aplicación genérica, sino que exigen una ponderación de los intereses en juego, el de acceso a la información pública, por un lado, y el protegido por la limitación de que se trate (vid. STS de 25 de enero de 2021 (rec. 6387/2019), FJ 4º.8), debiéndose interpretar los citados límites de forma restrictiva, a fin de no menoscabar el derecho de acceso, regulado de forma amplia en la Ley (vid. STS de 8 de abril de 2024 (rec. 681/2022), FJ 4º).

El principio de buena administración conduce también a una interpretación amplia y expansiva de este derecho constitucional, que conlleva una interpretación restrictiva de los límites oponibles al acceso a la información pública, con independencia de se exija su aplicación justificada y proporcionada, como examinaremos más adelante.

Este principio fue objeto de tratamiento en nuestra sentencia de 30 de abril de 2025 (rec. 1100/2022) en los términos que resumimos a continuación. Se infiere de los artículos 9.3 -proclama la garantía constitucional de la interdicción de la arbitrariedad de los poderes públicos-, 103 -la Administración Pública sirve con objetividad los intereses generales y actúa de acuerdo con los principios de eficacia, jerarquía, descentralización, desconcentración y coordinación, con sometimiento pleno a la ley y al Derecho- y 106 de la CE -los Tribunales controlan la legalidad de la actuación administrativa, así como el sometimiento de ésta a los fines que la justifican-.

Estos mandatos constitucionales tienen su reflejo en el artículo 3.1 de la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público, conforme al cual las Administraciones Públicas sirven con objetividad los intereses generales y actúan de acuerdo con los principios de eficacia, jerarquía, descentralización, desconcentración y coordinación, con sometimiento pleno a la Constitución, a la Ley y al Derecho, y deberán respetar en su actuación y relaciones, entre otros, los principios de buena fe, confianza legítima y lealtad institucional.

Además, el artículo 41 de la Carta de los Derechos Fundamentales de la Unión Europea que fue proclamada por el Parlamento Europeo, el Consejo de la Unión Europea y la Comisión Europea el 7 de diciembre de 2000 en Niza (DOUE núm. 83, de 30 de marzo de 2010), ha consagrado como un derecho fundamental de la Unión Europea el derecho a la buena administración.

Por último, el principio de buena administración ha sido objeto de tratamiento jurisprudencial, en diversas sentencias de esta Sala del Tribunal Supremo, de

las que destacamos la Sentencia núm. 1752/2022, de 23 de diciembre de 2022 (Rec. 1763/2021) que cita otros muchos precedentes jurisprudenciales, donde hemos enfatizado su efectividad, en la medida en que del mismo una serie de derechos de los ciudadanos con plasmación efectiva, y hemos precisado que no se trata, por tanto, de una mera fórmula vacía de contenido, sino que se impone a las Administraciones públicas de suerte que a dichos derechos sigue un correlativo elenco de deberes a estas exigibles, entre los que se encuentran, desde luego, la transparencia y el acceso a la información pública.

En esta línea de amplio reconocimiento del derecho de acceso, el artículo 16 de la LTAIBG prevé el acceso parcial a la información cuando resulte de aplicación alguno de los límites del anterior artículo 14, pero no afecte a la totalidad de la información, salvo que de ello resulte una información distorsionada o que carezca de sentido.

Por su parte, el artículo 15 de la LTAIBG contempla otra serie de límites derivados de la protección de datos personales que eventualmente pudiera contener la información solicitada, distinguiendo los supuestos en los que resultan involucrados datos especialmente protegidos (apartado 1º), datos meramente identificativos relacionados con la organización, funcionamiento o actividad del órgano (apartado 2º) y datos personales no especialmente protegidos (apartado 3º), y contemplando reglas específicas para cada uno de ellos, salvo que se haya efectuado una previa disociación de los datos personales que impida la identificación de las personas afectadas (apartado 4º), caso en el que no resultan aplicables aquellos límites.

IV. El USO DE LA IA POR LAS ADMINISTRACIONES PÚBLICAS. LAS DECISIONES ADMINISTRATIVAS AUTOMATIZADAS Y LAS EXIGENCIAS DE TRANSPARENCIA

IV.1.- La relevancia de la transparencia algorítmica ante el uso de sistemas de IA o aplicaciones de decisiones automatizadas por las Administraciones públicas.

Cuando nos referimos al empleo de sistemas de IA en la toma de decisiones, como es obvio, hemos de diferenciar las actuaciones basadas en el empleo de algoritmos de los sujetos privados y de las Administraciones Públicas, puesto que estas últimas están obligadas a cumplir las exigencias del Estado de derecho en el funcionamiento de los poderes públicos.

Nos centraremos en el uso de la IA por las Administraciones públicas para desarrollar la actividad administrativa y adoptar decisiones administrativas, dejando al margen la regulación administrativa del uso de la IA por sujetos

privados, cuya expansión en el funcionamiento de los poderes públicos pudiera conducirnos a la llamada algocracia -frente a la expresión burocracia weberiana-, entendida como un tipo particular de sistema de gobernanza organizado y estructurado en algoritmos programados por ordenador.

Algocracia: gobierno, administración o dirección de conductas mediante algoritmos.

La algocracia puede verse como una evolución o mutación digital de la burocracia. Si la burocracia clásica organiza el poder mediante normas, expedientes, jerarquías, formularios y funcionarios, la algocracia lo hace mediante datos, modelos predictivos, automatización, perfiles de riesgo, puntuaciones y sistemas de decisión algorítmica.

La algocracia no elimina la burocracia, sino que la reconfigura: sustituye parte de la decisión administrativa fundada en reglas escritas, jerarquía y expediente por decisiones o predecisiones automatizadas basadas en datos, patrones estadísticos y código informático.

La algocracia constituye una forma avanzada de racionalización burocrática en la que la decisión pública, sin abandonar necesariamente el marco formal de la Administración, queda condicionada por sistemas algorítmicos que clasifican, priorizan, predicen o incluso resuelven situaciones individuales. Ello desplaza el centro de gravedad desde la oficina, el expediente y el funcionario hacia el dato, el modelo y el código, generando nuevos problemas de transparencia, motivación, responsabilidad y control jurisdiccional.

En general, desde la perspectiva del Estado Social y Democrático de Derecho que proclama el artículo 1 de nuestra Constitución, se ha afirmado que el uso de la IA puede amenazar a la democracia como consecuencia de la divulgación de deepfakes o noticias falsas, la orquestación de campañas de desinformación o la adopción de decisiones bien equivocadas o bien caracterizadas por sesgos discriminatorios, con el consiguiente quebranto de los derechos fundamentales, como el principio de igualdad, pero resulta evidente que también puede llevar consigo ciertas ventajas para los propios sistemas políticos, al reforzar los procesos democráticos fomentando la participación de ciudadana en las decisiones administrativas mediante sistemas de IA o sirviendo como herramienta para la plena realización de los derechos fundamentales y mejorando la gestión administrativa en la prestación de servicios públicos y las políticas públicas al servirles de apoyo.

Desde un punto de vista técnico el empleo IA y las decisiones automatizadas en las Administraciones Públicas fomentan la eficacia del funcionamiento de los servicios públicos, al agilizar el procedimiento de decisión, reducir el personal asignado a determinadas tareas, favorecer el cumplimiento de las normas en la toma de decisiones y garantizar el principio de igualdad en su aplicación.

Sin embargo, también entraña riesgos importantes, especialmente cuando afecta a decisiones complejas, ante la eventualidad de que genere decisiones injustas o discriminatorias. Circunstancia, cuya singular gravedad se pone de manifiesto, cuando se recurre a sistemas de IA de caja negra, que son aquellos a los que se atribuye complejidad técnica, opacidad del contenido y restricciones en el acceso a su código fuente.

Así, la utilización de algoritmos y datos permite aumentar la capacidad de procesamiento de información, sin las limitaciones cognitivas y físicas de los seres humanos, proporciona información sobre la demanda de servicios públicos y permite adaptar la oferta de los mismos a las características de los usuarios, permite realizar predicciones tratando grandes cantidades de datos, evita los sesgos cognitivos humanos y las decisiones diferentes para supuesto análogos que requerirían la misma respuesta.

Por ello, se dice que contribuye a la realización de los principios constitucionales eficacia, imparcialidad y objetividad (artículo 103 CE) favoreciendo el principio de buena administración.

No obstante, también conlleva costes importantes, como los medioambientales por el impacto que el funcionamiento de estos sistemas genera, y riesgos no desdeñables, de entre los cuales destacaremos la discriminación algorítmica como consecuencia de sesgos y la opacidad

Estos riesgos se asocian, en primer lugar, al funcionamiento inadecuado, dado su impacto negativo a gran escala, que puede derivar de errores en la programación, de accidentes, con importante repercusión social o, simplemente, de errores de resultado fruto de su propio funcionamiento regular aunque imperfecto.

No debemos olvidar que los sistemas de aprendizaje automático o profundo realizan inducciones mediante métodos estadísticos aplicados a datos, por lo que las inducciones que generan sus resultados pueden no ofrecer un conocimiento cierto, en la medida que podrían existir datos no considerados por el sistema que pudieran hacer cambiar esa inducción o resultado.

Asimismo, al igual que se habla de la existencia de sesgos de ilusión causal en seres humanos que puede conducirles a encontrar relaciones de causalidad dónde no las hay, los sistemas de inteligencia artificial pueden emitir juicios sobre la base de escasa información, por ejemplo: (i) por falta de representatividad de la pregunta que se le formula en los datos manejados, lo puede provocar respuestas erróneas, (ii) por el empleo de datos del pasado que puede llevar a hacer predicciones basadas en la inercia histórica o el estatus quo, petrificando sus resultados, sin considerar variables inherentes a cambios futuros en la realidad a analizar, provocando la petrificación de los datos y la realidad.

Estos errores de resultado adquieren singular importancia ante la existencia de los llamados sesgos cognitivos humanos relacionados con el uso de sistemas de IA que infectan el sistema y los resultados de su uso. En este sentido se habla del sesgo de confirmación -resultado de la confianza excesiva en la información o decisiones adoptadas por una máquina, fruto de nuestra admiración por la ciencia-, el sesgo cognitivo de disponibilidad -lleva a los humanos actuar en función de la información más fácilmente disponible- aunque no sea la más

relevante o fiable- y el sesgo cognitivo de anclaje -lleva a decidir en función de la primera información recibida-

Naturalmente, estos sesgos afectan negativamente al sistema tanto cuando concurren en los programadores o las personas supervisoras del sistema, como cuando afectan a quién adopta la decisión final en el caso de decisiones con apoyo en sistemas de IA.

Además, junto a estos sesgos cognitivos humanos que afectan al funcionamiento del sistema, deben considerarse los riesgos de incorporación de sesgos en los propios datos empleados para nutrir los algoritmos, que se verían amplificados por el uso del sistema de IA.

Por tanto, la discriminación algorítmica se produce cuando los sistemas automatizados (modelos de aprendizaje automático, por ejemplo) reproducen o amplifican desigualdades preexistentes, ya sea por los sesgos contenidos en los datos de entrenamiento, por decisiones de diseño o por la falta de una adecuada supervisión humana, generando resultados discriminatorios e injustos. Este riesgo es particularmente grave cuando los modelos se aplican en ámbitos sensibles vinculados a la toma de decisiones.

En realidad, la IA hereda los sesgos de los datos con los que ha sido entrenada, lo que plantea la necesidad de diseñar e implementar sistemas que identifiquen y corrijan esos sesgos.

Ahora bien, los seres humanos también tienen sesgos, fruto de sus experiencias y el contexto en que desarrollan su actividad, y lo cierto es que sería posible diseñar sistemas de IA capaces de tomar decisiones que reflejen el sentir mayoritario de una comunidad, ajenas a los sesgos de las decisiones individuales y subjetivas. Desde esta perspectiva, serían capaces de decidir con mayor equidad y justicia, al manejar mayor cantidad de datos -experiencia-, carecer de influencias emocionales – emociones humanas o sentimientos- y tomar decisiones objetivas basadas en la lógica y los hechos.

Este planteamiento, conforme al cual los sistemas de IA, al verse desprovistos de emociones, podrían evitar los sesgos discriminatorios humanos que afectan negativamente a muchas de nuestras decisiones, encuentra como replica la idea de que la IA no solo no soluciona el problema de los sesgos humanos, sino que los ha ampliado.

Esto tiene lugar principalmente por dos vías: a) incorporación de sesgos humanos en los modelos de IA en todas sus fases de desarrollo (programación, entrenamiento con bases de datos sesgadas, contacto con usuarios sesgados en el mundo real) que se transmiten a muchas más personas, y b) el aprendizaje de sesgos de la IA por las personas que utilizan estos sistemas, cuestión que se está estudiando en el ámbito de la psicología experimental y que tiene por causa la confianza ciega en las herramientas de IA con las que se trabaja. En síntesis,

la IA aprende sesgos a partir de los humanos, los amplifica y los transmite a la siguiente generación de humanos, entrando así en un círculo vicioso del que podría ser difícil salir.

En verdad, el riesgo mayor de discriminación no se encuentra en la discriminación directa, donde una persona es tratada de forma menos favorable que otra en una situación similar, debido a una característica protegida (sexo, raza, orientación sexual o discapacidad, entre otras), sino en la discriminación indirecta que tiene lugar cuando una norma, criterio o práctica aparentemente neutra tiene un efecto perjudicial desproporcionado sobre un grupo protegido, sin que exista una justificación objetiva y razonable.

Naturalmente, resulta necesario tomar conciencia de estos riesgos y mitigarlos con la supervisión humana de la IA, ante la posibilidad de conculcar principios y derechos fundamentales a través del empleo de sistemas de IA, replicando o incluso intensificando desigualdades o discriminaciones ya existentes.

En el caso de estudio que incluye Eubanks (2021) en su libro se describen los problemas de un sistema inteligente utilizado en el estado de Pensilvania con el objetivo de predecir cuándo podría existir un caso de trato inadecuado a menores para enviar un trabajador social al domicilio y evaluar la situación. El modelo se retiró al poco tiempo porque se equivocaba en más del 70 % de las predicciones, penalizando a las familias más pobres. La explicación era clara: se había entrenado con datos de registros públicos y en Estados Unidos existe una mayor interacción de las familias con problemas económicos con las instituciones públicas.

En el ámbito de la justicia es bien conocido el comportamiento discriminatorio del sistema inteligente COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) (vid. Markou, 2017) para las personas afroamericanas. En algunos condados de Estados Unidos, los departamentos de libertad condicional utilizan evaluaciones de riesgo para sugerir al juez un plan de libertad condicional o un tratamiento apropiado para las personas que están siendo sentenciadas. Pero ser juzgado como inelegible para un tratamiento alternativo, particularmente durante una audiencia de sentencia, puede traducirse en encarcelamiento. Los acusados no suelen tener la oportunidad de impugnar sus evaluaciones. Los resultados generalmente se comparten con el abogado del acusado, pero los cálculos que transforman los datos subyacentes en una puntuación rara vez se revelan. El sistema inteligente COMPAS, comercializado por Nortpointe Inc., ha sido una de las herramientas de evaluación más utilizadas en los juzgados americanos entre 2001 y 2016, proporcionando una puntuación sobre riesgo de reincidencia de las personas juzgadas (sistema que analiza 137 parámetros a través de cuestionarios y del historial del delincuente, con información relativa al consumo de sustancias estupefacientes, el entorno familiar, los antecedentes penales o el grado de inserción social. Pero también entre esos parámetros estaría la pobreza, la raza, el nivel educativo y rasgos de la personalidad Utilizando esos criterios realizaría un scoring o puntuación predictiva del riesgo de reiteración delictiva de los presos y sujetos en libertad condicional, así como propondría otra serie de medidas de reinserción según el nivel adjudicado).

Existen diferentes ejemplos en la literatura que analizan el sesgo que produce COMPAS en las personas de origen afroamericano, encontrando casos en los que presenta un riesgo mucho mayor que para personas caucásicas y con tasas de error diferentes. En la clase de personas evaluadas con "mayor riesgo", pero que no reinciden, se tiene un error del 23.5 % para personas blancas frente a una tasa de fallo del 44.9 % para personas afroamericanas. Por el contrario, para la clase de personas con "menor riesgo", pero que vuelven a delinquir, la predicción falla en

un 47.7 % en personas blancas frente al 28.0 % en personas afroamericanas. De nuevo, estas estadísticas muestran un claro sesgo que tiene su origen en los datos que alimentan al sistema para su entrenamiento, con una población de origen afroamericano con mayor proporción en las cárceles y que no pudo identificarse al no disponer de un sistema de IA explicable.

Un caso que puso en tela de juicio el sistema fue el caso Loomis contra Wisconsin. En febrero de 2013, Eric Loomis fue condenado al conducir un coche usado en un tiroteo. Negó haber participado en el tiroteo, pero reconoció conducir el coche usado en él. Loomis fue condenado por los dos delitos más leves, resistencia a la autoridad y hurto de vehículo. Sin embargo, en la vista del juicio se consultó el sistema COMPAS y el tribunal fundó parte de la condena en dicha evaluación. La defensa recurrió por considerar vulnerado el derecho a un proceso debido porque el algoritmo del programa es secreto (lo que se denomina black box) y porque la sentencia se fundó entre otros factores en la raza del condenado, lo que sería discriminatorio. El Tribunal Supremo de Wisconsin desestimó el recurso considerándola no discriminatoria y fundada, en cuanto que la raza fue solo uno de los factores que tener en cuenta. De igual manera, a la vista de que el sistema utiliza información facilitada por el investigado, este podría negarse a contestar los cuestionarios y, así como reconoció la importancia de la individualización de las sentencias, sostuvo que se valoraron más pruebas y admitió que COMPAS solo compara el riesgo de reincidencia respecto a grupos sociales similares al del investigado.

En el ámbito laboral y empresarial de uso de algoritmos para analizar los perfiles de los candidatos, evaluar sus competencias y predecir su idoneidad para un puesto específico, destaca el caso de Amazon y su algoritmo de contratación como ejemplo paradigmático de cómo los sistemas de inteligencia artificial pueden perpetuar y amplificar sesgos existentes (BBC News Mundo, 2018; Dastin, 2018).

En 2014 Amazon inició un proyecto para desarrollar un sistema de IA que automatizara el proceso de selección de personal. La idea era crear una herramienta que pudiera revisar currículums y calificar a los candidatos de una a cinco estrellas, similar a cómo los clientes califican productos en su plataforma de comercio electrónico.

Sin embargo, para 2015, la compañía se dio cuenta de que el sistema mostraba un sesgo significativo contra las mujeres, especialmente para puestos técnicos como desarrollador de software. El problema radicaba en los datos de entrenamiento: el sistema había sido alimentado con currículums recibidos por Amazon durante los últimos diez años, un período en el que la industria tecnológica estaba dominada por hombres. Como resultado, el algoritmo "aprendió" que los candidatos masculinos eran preferibles. Llegó al punto de penalizar currículums que incluían la palabra "femenino" (como en "capitana de club de ajedrez femenino") y redujo la calificación de graduadas de dos universidades exclusivas para mujeres.

Otro riesgo del uso de los sistemas de IA se asocia a la opacidad, ligada a la complejidad técnica de la inteligencia artificial de aprendizaje automático o profundo qué hace muy difícil o imposible conocer cómo funcionan los algoritmos o qué datos se utilizan, dando lugar al concepto de cajas negras.

En general, la inteligencia artificial se caracteriza por la dificultad para conocer e interpretar su funcionamiento y explicar sus resultados. Esta opacidad puede derivar de factores de carácter técnico de la propia tecnología empleada y su funcionamiento (opacidad tecnólogica o intrinseca), vinculados a la naturaleza de los algoritmos, al desconocimiento de los datos utilizados para entrenar el sistema o a la reserva de las reglas que rigen el procesamiento de dichos datos. También puede tener su origen en factores jurídicos, ante el interés por preservar

otros derechos e intereses dignos de protección -datos de carácter personal, propiedad intelectual, secretos empresariales, seguridad de los algoritmos, confidencialidad en la toma de decisiones públicas, entre otros- limitándose la transparencia y el acceso a la información pública (opacidad jurídica). Por último, la opacidad puede derivar de factores organizativos como la resistencia de las Administraciones públicas a proporcionar información detallada sobre los sistemas de inteligencia artificial de que hace uso (opacidad organizativa).

Se dice que un sistema es una caja negra (black box) cuando su funcionamiento interno o lógica subyacente no se comprende adecuadamente o cuando no es posible explicar los resultados de salida. No obstante, actualmente, existen métodos para auditar las decisiones, que permiten desglosar los factores y datos que influyen en las decisiones de la IA, lo que facilita la comprensión de su proceso decisorio.

Ciertamente, éste es uno de los grandes riesgos derivados del uso de la inteligencia artificial, singularmente trascendente cuando tiene por causa la voluntad de las Administraciones públicas de ocultar el uso de los sistemas algorítmicos y los detalles acerca de su forma de funcionamiento. En términos generales, las aplicaciones de esta índole empleadas por las Administraciones Públicas no han sido aprobadas con carácter formal, al carecerse de una regulación procedimental *ad hoc*, y suelen ser desconocidas en la mayor parte de los casos, lo que genera opacidad, inseguridad jurídica y desconfianza en la actuación de los poderes públicos y, en consecuencia, menoscaba el principio de buena administración.

En este entorno adquiere especial relevancia el acceso a los algoritmos y sus códigos fuentes como manifestación del principio de transparencia y del derecho de acceso a la información pública, cuestión que fue abordada por la Sentencia del caso Bosco.

En todo caso, sea cual sea el alcance que quepa atribuir al principio de transparencia algorítmica, resulta imprescindible promover la transparencia y la explicabilidad algorítmica como principios esenciales para el funcionamiento de los sistemas de IA, mediante el impulso de sistemas de IA con la capacidad de explicar la justificación de las decisiones, caracterizar las fortalezas y debilidades de su proceso de toma de decisiones y transmitir una comprensión de cómo se comportarán en el futuro, con la finalidad de contribuir a superar el desconocimiento sobre el proceso de generación de resultados de los sistemas de IA.

En efecto, hoy es posible que los sistemas de IA ofrezcan a los usuarios una trazabilidad clara de sus resultados, informando sobre los factores o fórmulas empleados en la toma de decisiones, para justificarlas.

En Australia, Robodebt fue un polémico sistema automatizado de detección de fraudes del Gobierno australiano. El Departamento de Servicios Humanos de la Seguridad Social utilizó un

algoritmo de comparación de datos para comparar los ingresos registrados en el registro de Centrelink de un cliente con los datos históricos de ingresos informados por el empleador de la Oficina de Impuestos de Australia y emitió notificaciones automáticas de aumento y recuperación de deudas cada vez que se detectaban deudas sin ninguna supervisión humana. Se descubrió que el sistema había emitido avisos de cobro de deudas por valor de millones de dólares a miles de beneficiarios de asistencia social basándose en datos personales inexactos e información laboral errónea.

En 2020, tras la creciente presión pública y dos demandas perdidas, el Gobierno australiano declaró que Robodebt era ilegal, cerró el sistema y acordó condonar 470.000 deudas con reembolsos por un valor de 721 millones de dólares. En junio de 2021, un juez de un tribunal federal aprobó el acuerdo de una demanda colectiva de Robodebt por un valor de más de 1700 millones de dólares en beneficios económicos para aproximadamente 430 000 personas. El propio juez calificó el episodio de "El proceso ha puesto de manifiesto un capítulo vergonzoso en la administración del sistema de seguridad social de la Commonwealth y un fracaso masivo de la administración pública".

El Tribunal de Distrito de La Haya en 2020 dictó sentencia sobre el caso Systeem Risico Indicatie (SyRI), que era un instrumento legal utilizado por el Gobierno neerlandés para detectar diversas formas de fraude, incluidos en el cobro de beneficios sociales y asignaciones y fraude fiscal. El tribunal dictaminó que la legislación que regula el uso de SyRI viola la ley superior, esto es, que no se ajusta al artículo 8 del Convenio Europeo de Derechos Humanos (en adelante, CEDH), que protege el derecho al respeto de la vida privada y familiar, del domicilio y de la correspondencia, siendo que "los Países Bajos, como parte del CEDH, tienen una responsabilidad especial en la aplicación de nuevas tecnologías. El uso de "SyRI" no es suficientemente transparente y verificable" porque el sistema carecía de transparencia (ya que sus algoritmos no se publicaron y no se sometió a una auditoría técnica) y su selección de barrios desfavorecidos podría equivaler a una discriminación socioeconómica o por la condición de inmigrante. El sistema creaba informes de riesgo y se creó un "registro de denuncias de riesgo" con el fin de facilitar la información a las Administraciones y otros organismos participantes, al Ministerio Fiscal y a la Policía, siendo que, después de una investigación, pero las personas implicadas no eran informadas por separado sobre los informes de riesgo procesados en ese registro de denuncias. SyRI determinaba qué residentes locales merecían una investigación más a fondo La aplicación procesaba datos personales seudoanonimizados de todo tipo sobre el individuo, incluidos datos de endeudamiento, antecedentes penales, salud, etc., y de ella se extraían una serie de conclusiones sobre el nivel de riesgo de la persona que la Administración podía conservar durante dos años.

En este caso la negativa de remisión del expediente administrativo al tribunal contencioso-administrativo so pretexto de su complejidad técnica y su confidencialidad provocó que por ausencia de transparencia algorítmica y vulneración de la intimidad fuera anulada la herramienta de IA y prohibido su uso. Tras este caso el gobierno aprobó un procedimiento de control del empleo de algoritmos por las Administraciones Públicas y evaluación de su impacto en los derechos fundamentales.

Así, la transparencia se ha postulado como uno de los principios rectores de los sistemas de IA, pues disipar o reducir en la medida de lo posible la opacidad algorítmica es una exigencia del Estado democrático de derecho, consustancial al mismo, y aún más cuando el tratamiento realizado desemboca en decisiones públicas automatizadas.

Basta con reparar en el contenido de los considerandos 48 y 60 del RIA que identifican los derechos afectados, entre los que se destaca el principio a una

buena administración, para comprobar la preocupación del legislador europeo por tales riesgos.

Estos riesgos deben ser considerados y neutralizados para que el empleo de los sistemas de inteligencia artificial en el funcionamiento de las Administraciones Públicas posibilite la mejora de los servicios públicos, el reforzamiento de la democracia y el apoyo a las políticas públicas, lo cual exige una adecuada regulación del uso de estos sistemas en el ámbito del sector público.

En fin, de lo que no cabe duda es que la inteligencia artificial empleada por las Administraciones públicas ha de quedar sometida al principio de legalidad propio del Estado de derecho, así como al principio de seguridad jurídica, garantizándose la aplicación de las normas que contribuyan a la existencia de una inteligencia artificial fiable, respetuosa con los derechos fundamentales (protección de datos personales, igualdad, no discriminación, etc.) y otras exigencias de una sociedad democrática, como el principio de buena administración.

Además, la transparencia algorítmica debe someterse a los principios que regulan el acceso a la información pública, por lo que, a priori, los algoritmos empleados por las Administraciones públicas deberían quedar sometidos a dicho acceso, sin perjuicio de los límites que establece la LTAIPBG y su alcance en cada concreto caso.

Es más, la transparencia y el acceso a la información pública que aborda la LTAIPBG implica exigencias de “transparencia algorítmica” dirigidas a garantizar la explicabilidad de los sistemas de IA, asegurando la comprensibilidad de su funcionamiento mediante la descripción del diseño del algoritmo en lenguaje natural y así lo ha afirmado nuestra jurisprudencia en la STS caso BOSCO.

Verdaderamente, la doctrina científica se aproxima a estos conceptos -transparencia algorítmica y explicabilidad algoritmica- configurándolos como conceptos técnicos íntimamente relacionados, pero con diferente significado y trascendencia.

Así, se dice que la transparencia algorítmica se refiere a la disponibilidad de información sobre los componentes, procesos y objetivos de un sistema algorítmico, con la finalidad de permitir su supervisión por parte de entidades reguladoras, auditores o incluso por la ciudadanía. Su cumplimiento supone exigencias de documentación adecuada, carácter explícito de sus finalidades, trazabilidad de los datos de entrenamiento e identificación de los responsables del sistema. Su finalidad es conocer los datos que un algoritmo utiliza, cómo lo usa, quiénes los emplean, para qué fin y cómo se toman las decisiones, lo que puede abarcar tanto el diseño de la plataforma y los mecanismos algorítmicos como la lógica subyacente del sistema de software, todo lo cual permite verificar que funciona adecuadamente -sin sesgos ni errores-.

Por otra parte, se dice que la explicabilidad algorítmica alude a la capacidad del sistema para ofrecer razones inteligibles y comprensibles sobre los procesos técnicos de un sistema de IA y las decisiones concretas asociadas al mismo, especialmente desde la perspectiva del sujeto afectado. Su finalidad es que el sujeto afectado pueda entender porque se ha adoptado una decisión concreta por el sistema, qué factores fueron determinantes para ello y cómo puede ser impugnada o revisada.

Con arreglo a esta distinción conceptual, la transparencia algorítmica estaría orientada, fundamentalmente, al control institucional y la gobernanza, mientras que la explicabilidad algorítmica se configuraría como una garantía de los derechos fundamentales de las personas afectadas, no como una mera herramienta de auditoría. Por ello, se dice que la transparencia es un requisito estructural *ex ante*, mientras que la explicabilidad constituye una garantía funcional *ex post*, orientada a la satisfacción del derecho a comprender y a un trato justo en contextos de automatización.

Observamos como la transparencia algorítmica es una exigencia del derecho europeo, recogida en el Reglamento General de Protección de Datos de carácter personal (artículos 5.1 punto a, 13 y 14, en general en el tratamiento de datos, y 22, respecto de decisiones automatizadas) y el Reglamento de Inteligencia Artificial (artículos 13 y 52), y la explicabilidad algorítmica se refleja en la configuración de los derechos digitales (artículos 8 y 47 de la carta de los derechos fundamentales de la Unión Europea), si bien también tiene reflejo cómo exigencia normativa, junto con la transparencia, en el Reglamento de Inteligencia Artificial.

Sin embargo, pese a esa útil distinción técnica entre ambos conceptos, tanto por su naturaleza, su finalidad y sus destinatarios, sostenida por la doctrina científica, lo cierto es que ambas dimensiones –transparencia y explicabilidad- se encuentran íntimamente conectadas y resultan indispensables para articular un verdadero régimen de rendición de cuentas algorítmica como instrumentos complementarios en la arquitectura del derecho de la inteligencia artificial.

Por ello, en la doctrina científica tampoco es inusual que ambos conceptos, transparencia y explicabilidad se integren en un solo concepto de transparencia algorítmica, entendido en sentido amplio, que comprendería tanto su dimensión estructural como la funcional. Desde esta perspectiva, se dice que la transparencia algorítmica está al servicio de su explicabilidad y se relaciona con la capacidad de los ciudadanos para entender las operaciones que realizan los algoritmos, por lo tanto, está vinculada al conocimiento de las personas al interactuar con los algoritmos, independientemente de que una organización los implemente.

Así, se dice que «transparencia» es un concepto inclusivo que recoge exigencias como trazabilidad, explicabilidad, comunicación, donde se insertan derechos

como la «interpretabilidad», la «auditabilidad», la «testabilidad», «comprobabilidad», «verificabilidad» y «replicabilidad» del mismo, entendidos todos ellos como atributos exigibles a un sistema de IA «transparente».

En sintonía con esta última reflexión, en adelante, emplearemos el concepto de transparencia algorítmica en sentido amplio, comprensivo de la explicabilidad algorítmica.

Pues bien, realizadas estas consideraciones conceptuales, debemos enfatizar que la aplicación del derecho de transparencia y acceso a la información pública -en la interpretación jurisprudencial que se ha hecho de su configuración constitucional y su regulación legal- adquiere singular trascendencia cuando se proyecta sobre la actividad automatizada de la Administración a través de aplicaciones informáticas, con o sin uso de IA, puesto que, como ha apuntado en diversas ocasiones el Tribunal Constitucional, *«[h]abida cuenta de que nuestro texto constitucional no consagra derechos meramente teóricos o ilusorios, sino reales y efectivos (STC 12/1994, de 17 de enero, FJ 6), se hace imprescindible asegurar su protección no sólo frente a las injerencias tradicionales, sino también frente a los riesgos que puedan surgir en una sociedad tecnológicamente avanzada»* (vid. STC 16/2004, de 23 de febrero, FJ 3º).

En efecto, la configuración del derecho de acceso a la información pública que hemos expuesto adquiere especial relevancia ante los riesgos que entraña el uso de las nuevas tecnologías en el ejercicio de las potestades públicas o la prestación de servicios públicos, como ocurre con el empleo de sistemas informáticos de toma de decisiones automatizadas en la actividad de las Administraciones públicas, especialmente, cuando, como ocurría en la STS caso BOSCO, tienen por objeto el reconocimiento de derechos subjetivos de los ciudadanos (y, más aún, cuando se trata de derechos de carácter social, atribuibles a los ciudadanos más desfavorecidos o necesitados de protección)

Con la finalidad de alcanzar una mayor eficiencia en la gestión de los recursos públicos las Administraciones Públicas utilizan frecuentemente los sistemas de inteligencia artificial en distintas áreas de su actividad, tanto para recopilar un volumen ingente de datos y analizarlos con el fin de diseñar sus políticas públicas, como para adoptar de forma automatizada decisiones que repercuten de forma directa sobre los derechos individuales de los ciudadanos: lucha contra el fraude fiscal, selección de beneficiarios de servicios y ayudas sociales, ámbito sanitario (diagnósticos y gestión de hospitales y de listas espera), asignación de becas académicas, evaluación del rendimiento de empleados públicos, contratación administrativa, seguridad pública (sistemas de predicción de riesgo desplegados en relación con la comisión de actos delictivos o infracciones administrativas, como Viogen o Riscanvi), entre otras materias.

El problema es que esta proliferación exponencial de sistemas de decisión automatizada en el sector público se ha producido en un vacío regulatorio preocupante, pues el artículo 41 de la LRJSP se limitaba a definir las

«actuaciones administrativas automatizadas» sin establecer garantías específicas de transparencia algorítmica, la normativa sectorial guardaba silencio sobre la obligación de publicar o facilitar acceso a los códigos fuente y la LTAIBG no contemplaba específicamente la problemática en relación al objeto de la propia ley en relación a los algoritmos públicos.

Ello ha generado una evidente opacidad en la práctica administrativa sobre este particular, al no publicar las características de estos sistemas, no explicar su lógica decisoria, no permitir auditorías independientes y no facilitar acceso ciudadano a sus códigos fuente, lo que generaba consecuencias intolerables desde la perspectiva del Estado de Derecho, ante la imposibilidad de control de legalidad del algoritmo, la indefensión material de los afectados por desconocimiento de la motivación de las decisiones administrativas, imposibilidad de detección de sesgos discriminatorios en los algoritmos y debilitamiento de la rendición de cuentas de las Administraciones Públicas y del control democrático sobre las mismas.

Con el objetivo de paliar o remediar la opacidad de los sistemas de IA empleados por las Administraciones públicas y canalizar la transparencia algorítmica se han puesto en marcha iniciativas de diferente naturaleza. Entre ellas, destaca la creación de registros de algoritmos y bases de datos que recogen información sobre los sistemas algorítmicos utilizados por las Administraciones Públicas (en comunidades autónomas como la Generalitat de Cataluña, en países como Canadá, Chile, Países Bajos y Reino Unido y en ciudades como Ámsterdam, Barcelona, Bruselas, Eindhoven, Mannheim, Rotterdam y Sofía).

Otro mecanismo para promover la transparencia algorítmica es la difusión del código fuente de los algoritmos, aunque, a veces, no resulta suficiente para cumplir tal finalidad, especialmente, cuando nos enfrentamos a modelos de aprendizaje automático o cuando los destinatarios de la información carecen de la formación técnica adecuada para su comprensión.

Asimismo, la transparencia de los datos utilizados para el entrenamiento o la evaluación de los sistemas de IA, que resulta necesario para conocer el funcionamiento de los algoritmos y poder supervisarlo, se puede ver favorecido por la difusión por las Administraciones Públicas de datos abiertos y reutilizables, así como la promoción de su uso por los sistemas de IA (véase el artículo 5.3 de la Ley 37/2007, de 16 de noviembre, de reutilización de la información en el sector público, a la que hace referencia el artículo 5.4 de la LTAIPBG que promueve su reutilización), aunque lo cierto es que no es frecuente que los portales de datos abiertos de las Administraciones Públicas difundan los datos utilizados por los algoritmos

Desde luego, no cabe cuestionar la conveniencia de que las Administraciones públicas recurran a sistemas de toma de decisiones automatizadas para el eficaz desempeño de sus funciones o la adecuada prestación de servicios públicos. No

obstante, ello debe conllevar exigencias de transparencia de los procesos informáticos seguidos en dichas actuaciones, con el objeto de proporcionar a los ciudadanos la información necesaria para su comprensión y el conocimiento de las características básicas de su funcionamiento, lo que puede requerir el acceso a su código fuente.

Aunque se ha afirmado que el derecho a la transparencia algorítmica en las actuaciones de las Administraciones Públicas, derivado del derecho de acceso a la información pública, puede ser encuadrado entre los llamados derechos de cuarta generación vinculados con las tecnologías de la información y de la comunicación, donde se han abierto paso una amplia variedad de derechos digitales (protección de derechos de carácter personal, a la identidad en el entorno digital, acceso a internet, a la ciberseguridad, a la desconexión digital, etc.), lo cierto es que este derecho no se encuentra positivizado en nuestro ordenamiento jurídico, si bien encontramos referencias al mismo en algunos instrumentos internacionales.

Mas allá de la constatación de esta realidad de ausencia de positivización del derecho de transparencia algorítmica, la configuración del derecho de acceso a la información pública que hemos expuesto adquiere especial relevancia ante los riesgos que entraña el uso de las nuevas tecnologías en el ejercicio de las potestades públicas o la prestación de servicios públicos, como ocurre con el empleo de sistemas informáticos de toma de decisiones automatizadas en la actividad de las Administraciones públicas, especialmente, cuando, como acontecía en la sentencia del caso Bosco, tienen por objeto el reconocimiento de derechos subjetivos de los ciudadanos y, más aún, cuando se trata de derechos de carácter social, atribuibles a los ciudadanos más desfavorecidos o necesitados de protección, donde el Tribunal Supremo consideró que el caso era merecedor de un estándar de transparencia reforzada,

Cuando las Administraciones Públicas hacen uso de sistemas informáticos de toma de decisiones automatizadas en el ejercicio de las potestades públicas, con afectación de los derechos de los ciudadanos, el acceso a su código fuente es uno de los mecanismos a través de los cuales se garantiza la transparencia algorítmica que demanda el pleno ejercicio del derecho de acceso a la información pública. No obstante, debe reconocerse que la autorización de ese acceso puede entrañar riesgos para otros derechos o intereses dignos de protección, que deben ser considerados y ponderados, bajo el marco legal de los límites al derecho de acceso a la información pública y maximizando este acceso.

En estos casos la transparencia de las aplicaciones informáticas o del proceso tecnológico seguido por el sistema informático adquiere singular relevancia para garantizar el adecuado control de la gestión pública, al brindar a la ciudadanía la información necesaria acerca del proceso seguido en la toma de decisiones para su comprensión, así como para comprobar su adecuación a las normas cuya aplicación debe regir su funcionamiento.

Así es, la explicabilidad de las aplicaciones informáticas, así como de los algoritmos que las sustentan, utilizadas por las Administraciones públicas, es objeto de una creciente demanda ciudadana, como condición inexcusable para preservar la rendición de cuentas y la fiscalización de las decisiones de los poderes públicos y, en último término, como garantía efectiva frente a la arbitrariedad o los sesgos discriminatorios en la toma de decisiones total o parcialmente automatizadas (el acceso al código fuente posibilita la comprobación de la conformidad del sistema algorítmico con las previsiones normativas que debe aplicar).

En fin, en estos casos la transparencia de las aplicaciones informáticas o del proceso tecnológico seguido por el sistema informático adquiere singular relevancia para garantizar el adecuado control de la gestión pública, al brindar a la ciudadanía la información necesaria acerca del proceso seguido en la toma de decisiones para su comprensión, así como para comprobar su adecuación a las normas cuya aplicación debe regir su funcionamiento.

Así es, en este escenario de desarrollo digital adquiere especial trascendencia, entre los derechos digitales de los ciudadanos, el principio de transparencia mediante el cual se considera que todo el mundo debería tener acceso a una información comprensible y precisa sobre los sistemas tecnológicos que afectan a su vida, así como la capacidad de poner en duda y cambiar aquellos que resulten injustos, parciales o discriminatorios (declaración de la coalición de ciudades por los derechos digitales).

Pues bien, cabe afirmar que cumplir con tal exigencia de transparencia de las aplicaciones informáticas o los sistemas de IA comprende: (i) que los individuos sean informados de que están interactuando con un sistema de esos caracteres, así como de sus capacidades y limitaciones, lo que permitirá que sean conscientes de los riesgos que supone su empleo; (ii) aportar a los ciudadanos información suficiente de estos sistemas y la forma en que adoptan sus decisiones, dada la incidencia que ello tiene sobre su derecho de defensa frente a las decisiones o resultados que pudieran contravenir el ordenamiento jurídico, facilitando su impugnación eficaz, y (iii) que la información facilitada al usuario resulte comprensible para él, permitiendo la comprensión de la lógica y los patrones empleados por el sistema para alcanzar los resultados obtenidos.

Verdaderamente, la explicabilidad de los sistemas de IA o las aplicaciones informáticas, los empleen o no, constituye el primer nivel de garantía de la transparencia algorítmica, cuyo propósito o finalidad es asegurar que las decisiones derivadas de sistemas automatizados e inteligentes sean comprensibles y auditables, proporcionando explicaciones significativas en un lenguaje claro y comprensible a los ciudadanos que permita conocer las soluciones tecnológicas implementadas por las administraciones, los datos de que se nutre, el proceso de toma de decisión, la existencia de supervisión humana sobre su funcionamiento y su alcance, etc. Información, en definitiva, que posibilite el control público sobre la racionalidad de los sistemas de inteligencia artificial o automatizados sin requerir el acceso directo al código fuente

En relación con estas exigencias, cuando las Administraciones Públicas hacen uso de sistemas informáticos de toma de decisiones automatizadas en el ejercicio de las potestades públicas, con o sin uso de inteligencia artificial, con afectación de los derechos de los ciudadanos, se suscita el debate de si ello implica, en virtud del respeto al principio de transparencia y acceso a la información pública, el acceso público a los algoritmos y los códigos fuente utilizados en la toma de esas decisiones administrativas.

Como es lógico, este debate nos conduce al terreno de los límites del acceso a la información pública, especialmente, los límites de propiedad intelectual, secreto profesional y seguridad, y con ello a la reflexión de que en los sistemas más complejos, los sistemas de IA, no siempre el conocimiento del código fuente resulta suficiente para conocer los patrones que llevan a la máquina a obtener unos determinados resultados, lo que resulta frecuente en los llamados modelos de caja negra que no permiten descifrar el razonamiento empleado por el sistema para alcanzar los resultados obtenidos.

Como se ha afirmado, los algoritmos y códigos fuente utilizados por las Administraciones Públicas en la toma de decisiones automatizadas constituyen información pública a los efectos de la legislación reguladora de la transparencia, por lo que la divulgación, a través de su publicación en los portales de transparencia, en su caso, con los límites a que hace referencia el artículo 5.3 de LTAIBG, o el acceso directo a los mismos puede constituir una fórmula idónea para garantizar el principio de transparencia, como condición indispensable para preservar la rendición de cuentas y la fiscalización de las decisiones de los poderes públicos y, en definitiva, como garantía efectiva frente a la arbitrariedad o los sesgos discriminatorios en la toma de decisiones total o parcialmente automatizadas.

En verdad, el acceso a su código fuente es uno de los mecanismos a través de los cuales se garantiza la transparencia algorítmica que demanda el pleno ejercicio del derecho a la información pública, pero no el único, ni, en ocasiones, el más adecuado. Además, debe reconocerse que la autorización de ese acceso puede entrañar riesgos para otros derechos o intereses dignos de protección, que deben ser considerados y ponderados, bajo el marco legal de los límites al derecho de acceso a la información pública.

Desde luego, no puede negarse que no siempre el mecanismo más útil para garantizar la transparencia algorítmica es la publicación del algoritmo, y que puede bastar con facilitar la explicabilidad de los sistemas de IA, garantizando la comprensibilidad de su funcionamiento mediante la descripción del diseño del algoritmo en lenguaje natural.

Por último, destacamos que el acceso a los algoritmos de los sistemas electrónicos empleados por las Administraciones Públicas, ya sea como apoyo a la decisión o en la elaboración de decisiones electrónicas automatizadas, se ha relacionado con el derecho de acceso al expediente administrativo de los interesados reconocido en el artículo 35 de la Ley 39/2015, de 1 de octubre, del Procedimiento Administrativo Común de las Administraciones Públicas (LPAC),

pretendiéndose que este acceso comprenda los algoritmos. Bajo este presupuesto, ante una eventual impugnación judicial de la decisión administrativa, tal acceso sería presupuesto del derecho de defensa y a un proceso debido, en íntima relación con principios diversos, como los de igualdad de partes, contradicción, prohibición de indefensión y motivación de resoluciones judiciales.

Y también se ha relacionado con el derecho a la libertad de información, considerándose que el derecho a conocer los algoritmos empleados en las decisiones adoptadas por las Administraciones Públicas formaría parte del contenido de la libertad de recibir información veraz sobre herramientas tecnológicas en poder de las administraciones, cuyo funcionamiento incide en los derechos de los ciudadanos.

IV.2.- Marco normativo interno.

En el examen de nuestra normativa sobre las actuaciones administrativas automatizadas dejaremos a un lado las exigencias sobre transparencia que se contienen en los reglamentos europeos, de las que nos ocuparemos más adelante, aplicables solo cuando afecten a derechos de carácter personal o a empleo de sistemas de IA.

La actuación administrativa automatizada que proporciona una decisión automatizada es una solución tecnológica que mediante el tratamiento de datos proporciona la consecución o el cumplimiento de un determinado objetivo sin intervención humana.

Descendiendo al **ámbito de las decisiones administrativas automatizadas** cabe reseñar que el principio de transparencia en el ámbito de la administración pública aparece reflejado en la ley 39/2015 (artículos 3, 71, 29 y 132) en la ley 40/2015 (artículos 3, 38, 81 y 112), pero no se hace alusión a la transparencia algorítmica.

La actuación administrativa automatizada se reguló por primera vez en nuestro ordenamiento jurídico mediante la Ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los servicios público, que siguió el modelo ya incorporado en la Ley 58/2003, de 17 de diciembre, General Tributaria (artículo 96) y ha sido replicada en mayor o menor medida por la normativa de las distintas comunidades autónomas.

Posteriormente, esta regulación fue incorporada con carácter básico en el artículo 41.1 de la ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público (LRJSP), y desarrollada por el 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.

El **artículo 41 de la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público,** se limita a definir las «actuaciones administrativas automatizadas» sin establecer garantías específicas de transparencia algorítmica, la normativa sectorial guardaba silencio sobre la obligación de publicar o facilitar acceso a los códigos fuente y la Ley 19/2013 de Transparencia no contempla específicamente la problemática en relación al objeto de la propia ley en relación a los algoritmos públicos.

«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».

El precepto en relación con estas actuaciones exige que se establezca con carácter previo 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.

Igualmente, esa misma norma exige que se indique el órgano que debe ser considerado responsable a efectos de impugnación.

La actuación administrativa automatizada se refiere, por tanto, a cualquier actuación realizada íntegramente a través de medios electrónicos por una administración pública en el seno procedimiento administrativo, sin intervención directa de un empleado público, si bien la actuación se atribuirá a un concreto órgano administrativo -que ejerce la potestad y tiene encomendada la competencia- y el sistema electrónico de información y su código fuente quedará sometido a supervisión humana.

Asimismo, el **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**, se limita a establecer que la Sede Electrónica debe publicar la relación actualizada de las actuaciones administrativas automatizadas vinculadas a los servicios, procedimientos y trámites y, en particular, la descripción de su diseño y funcionamiento, los mecanismos de rendición de cuentas y transparencia, así como los datos utilizados en su configuración y aprendizaje, desarrollando de este modo el artículo 41 de la Ley 40/2015 (véase el artículo 11).

También deben tenerse en cuenta el **artículo 96 de la Ley General Tributaria** sobre “Utilización de tecnologías informáticas y telemáticas” y el **artículo 130 del Texto refundido de la Ley General de la Seguridad Social**, aprobado por Real Decreto Legislativo 8/2015, de 30 de octubre, sobre “Tramitación electrónica de procedimientos en materia de Seguridad Social”, que presentan similar contenido al artículo 41 de la LRJSP.

Por lo que se refiere a la protección de datos personales, ha de tenerse presente la **Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales,** que en este particular se limita a reiterar lo que establece el RGPD.

Por último, la **Ley 15/2022, de 12 de julio, Integral para la igualdad de trato y la no discriminación** contiene un artículo referido a los sistemas de IA, concretamente su **artículo 23** sobre "Inteligencia Artificial y mecanismos de toma de decisión automatizados

Ninguna de estas normas se refiere a la transparencia algorítmica permitiendo el acceso al código fuente de los algoritmos empleados por las Administraciones pública, si bien el artículo 23 de la Ley 15/2022, de 12 de julio, Integral para la igualdad de trato y la no discriminación se refiere a la transparencia y a la Carta de Derechos Digitales.

No podemos dejar de mencionar la **Carta Española de Derechos Digitales**, de 14 de julio de 2021, que pese a carecer de fuerza normativa vinculante, recoge 26 derechos y, concretamente, en su apartado XVIII.6, dispone

«Derechos digitales de la ciudadanía en sus relaciones con las Administraciones Públicas

(…)

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

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

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

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

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

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

En el apartado XXV.2.b se establece que:

«Derechos ante la inteligencia artificial.

(..)

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

(…)

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

Esta Carta fue el fruto del trabajo de un grupo de expertos a iniciativa de la Secretaría de Estado de Digitalización e IA, y su naturaleza es la de mera declaración política de intenciones, no es una norma jurídica vinculant

En el **ámbito autonómico** se han aprobado también algunas leyes y disposiciones reglamentarias sobre inteligencia artificial, de las que destacamos las siguientes:

- El Decreto Ley extremeño 2/2023, de 8 de marzo, de medidas urgentes de impulso a la inteligencia artificial en Extremadura (incorpora la Carta de Derechos Digitales en su artículo 11.1 -artículos 11 y 12-)
- La Ley 1/2022, de 13 de abril, de Transparencia y Buen Gobierno de la Comunitat Valenciana, que impone la publicación de la relación de sistemas algorítmicos o de inteligencia artificial que tengan impacto en los procedimientos administrativos o la prestación de los servicios públicos con la descripción de manera comprensible de su diseño y funcionamiento, el nivel de riesgo que implican y el punto de contacto al que poder dirigirse en cada caso, de acuerdo con los principios de transparencia y explicabilidad.
- La Ley 6/2024, de 5 de diciembre, de simplificación administrativa de la Comunidad Valenciana (artículo 43).
- La ley 7/2024, de 11 de diciembre, de medidas urgentes de simplificación y racionalización administrativas de las Administraciones Públicas de les Illes Baleares (artículo 67).
- Ley 2/2025, de 2 de abril, de simplificación administrativa de Cantabria (artículo 50)
- Ley 2/2025, de 2 de abril, para el desarrollo y el impulso de la inteligencia artificial en Galicia (artículo 1, 5 y ss, 22 y ss y 53).
- Ley 4/2025, de 11 de julio, de Simplificación, Agilización y Digitalización Administrativa d Castilla-La Mancha (arts. 45 y ss).
- Decreto 76/2020, de 4 de agosto, de Administración digital de Cataluña (art. 54.2).
- Decreto del País Vasco 21/2012, de 21 de febrero de Administración electrónica.

Por tanto, las actuaciones administrativas automatizadas se implantaron hace ya tiempo en nuestras Administraciones públicas. Sin embargo, solo recientemente y con motivo del desarrollo de los sistemas de inteligencia artificial, se están llevando a cabo actividades administrativas automatizadas de cierta complejidad pues hasta ahora se empleaban solo para tareas sencillas, como la expedición de recibos, registros electrónicos, comprobación automática de datos de solicitudes, impulso automático de los procedimientos, publicación de notificaciones y anuncios en el BOE o el intercambio automático de datos entre Administraciones públicas –trámites regulados en los artículos 16, 66, 71 y disposición adicional tercera de la LPAC y en el artículo 44 de la LRJSP-.

No cabe duda de que resulta necesaria una regulación del uso de la inteligencia artificial en la Administración pública, fundamentalmente, a través de la reforma de nuestra legislación básica, Leyes 39/2015 y 40/2015, en materia de procedimiento administrativo común y régimen jurídico del sector público

(artículo 149. 1.18 CE *"procedimiento administrativo común y las bases del régimen jurídico de las Administraciones públicas"* y 18.4 CE *"La ley limitará el uso de la informática para garantizar el honor y la intimidad personal y familiar de los ciudadanos y el pleno ejercicio de sus derechos"*), sin perjuicio de la regulación que en ejercicio de sus competencias pudiera estar pudieran establecer cada una de las comunidades autónomas y el correspondiente desarrollo reglamentario.

La legislación básica estatal debería: (i) comprender los derechos de los ciudadanos ante el uso de la inteligencia artificial en las Administraciones Públicas y las obligaciones de derivadas de esos derechos; (ii) actualizar los principios de actuación de las Administraciones Públicas vinculados al uso de la inteligencia artificial, y (iii) ocuparse tanto de la regulación del procedimiento de automatización de la actividad administrativa, como del procedimiento mediante el cual tiene lugar la decisión administrativa automatizada, así como el empleo de estos sistemas de IA como instrumento de apoyo a la toma de decisiones administrativas.

En relación con esto último, se debería concretar qué tipo de actuaciones puede llevar a cabo la Administración pública a través de sistemas completamente automatizados, concretamente, si debería limitarse a actos meramente reglados o puede abarcar también actos de naturaleza discrecional, así como, en el primer caso, sí podría implicar actividades que supongan juicios de valor o no –cuestión sobre la que se ha pronunciado alguna normativa autonómica, como la ley 7/2014, de 11 de diciembre, de medidas urgentes de simplificación y racionalización administrativas de las Administraciones Públicas de Les Illes Baleares que prohíben las actuaciones administrativas automatizadas que supongan juicios de valor, así como la legislación de algunos de los países de nuestro entorno, como Alemania, cuya Ley de procedimiento administrativo fue modificada en 2017 para introducir una disposición relativa a la decisión administrativa completamente automatizada (párrafo 35), donde se establece la prohibición de dicha clase de actuaciones cuando exista discrecionalidad o margen de apreciación-

En relación con estas necesidades regulatorias adquiere singular interés el intenso debate doctrinal existente sobre si el código fuente, en la medida que suponga la traducción al lenguaje informático de una norma jurídica, debe ser considerado un reglamento, lo que conllevaría que para su elaboración y aprobación se siguiera un procedimiento reglamentario, sujeto a información pública y publicación y susceptible de los cauces de impugnación judicial propios de las disposiciones de carácter general -directo e indirecto-.

Por el contrario, en caso de no suponer la traducción al lenguaje informático de una norma jurídica, se trataría de un acto administrativo, que también podría ser ser sometido a información pública, ex artículo 83 de la LPAC, y publicación, ex artículo 45 LPAC.

También resulta de notable interés la polémica existente acerca de si debiera establecerse una reserva de humanidad para determinadas actuaciones administrativas, es decir, que quedarán vedadas a la automatización. Parece que exigencias de buena administración imponen ese límite para las actuaciones administrativas discrecionales, con la justificación de que el ejercicio de la discrecionalidad implica un juicio cognitivo, emocional y volitivo encaminado obtener el mejor resultado posible para la satisfacción de los intereses generales, que en la actualidad solo los seres humanos pueden llevar a cabo. La ausencia de empatía en las máquinas y su incapacidad para desarrollar juicios o inferencias abductivas, a diferencia de lo que ocurre con los seres humanos, impide que puedan ejercer adecuadamente las potestades discrecionales.

La Carta española de Derechos Digitales de 2021 reconoce el derecho a "[q]ue la adopción de decisiones discrecionales quede reservada a personas, salvo que normativamente se prevea la adopción de decisiones automatizadas con garantías adecuadas" (apartado XVIII, 6, d)).

El Reglamento de IA, por su parte, establece también reserva de humanidad en tanto que prohíbe el uso de la IA en ciertos casos (artículo 5) y exige supervisión humana y obligaciones de transparencia para usuarios y personas afectadas en los sistemas de alto riesgo (artículo 14).

V.3.- Marco normativo europeo.

Al respecto, no debe olvidarse las normas relativas a los derechos fundamentales y a las libertades que la Constitución reconoce se deben interpretar de conformidad con la Declaración Universal de Derechos Humanos y los tratados y acuerdos internacionales sobre las mismas materias ratificados por España, ex artículo 10.2 de la Constitución, y que los tratados internacionales válidamente celebrados, una vez publicados oficialmente en España, formarán parte del ordenamiento interno, ex artículo 96.1 de la Constitución.

En el marco de la normativa europea encontramos dos normas de aplicación sectorial o parcial -protección de datos de carácter personal y sistemas de IA-, pero ninguna de ellas desarrolla la transparencia algorítmica hasta el punto de exigir acceso al código fuente de los algoritmos empleados por las Administraciones públicas.

1.- Por un lado, el **Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos** y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos) (RGPD) -al igual que Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales-, establece un régimen específico de transparencia cuando se

trata de toma de decisiones plenamente automatizadas en el marco de lo previsto en su artículo 22, al margen de las obligaciones de transparencia aplicables a todo tratamiento de datos (artículos 5, 12, 13, 14 y 15).

El artículo 22 del RGPD se refiere a las "Decisiones individuales automatizadas, incluida la elaboración de perfiles":

«1. Todo interesado tendrá derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en él o le afecte significativamente de modo similar.

2. El apartado 1 no se aplicará si la decisión:

a) es necesaria para la celebración o la ejecución de un contrato entre el interesado y un responsable del tratamiento;

b) está autorizada 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, o

c) se basa en el consentimiento explícito del interesado.

3. En los casos a que se refiere el apartado 2, letras a) y c), el responsable del tratamiento adoptará las medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado, como mínimo el derecho a obtener intervención humana por parte del responsable, a expresar su punto de vista y a impugnar la decisión.

4. Las decisiones a que se refiere el apartado 2 no se basarán en las categorías especiales de datos personales contempladas en el artículo 9, apartado 1, salvo que se aplique el artículo 9, apartado 2, letra a) o g), y se hayan tomado medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado.»

Por tanto, el precepto establece una prohibición para las decisiones individuales automatizadas, incluyendo expresamente la elaboración de perfiles, siempre que concurran dos condiciones: la falta de intervención humana y la producción de efectos jurídicos o afectación significativa a un interesado.

No obstante, esta prohibición admite excepciones cuando se cumplan las condiciones contempladas en el apartado segundo del precepto, entre las que se encuentra la existencia de habilitación normativa para que las Administraciones Públicas puedan adoptar decisiones basadas exclusivamente en el tratamiento automatizado de datos personales, que operará, aunque la decisión afecte significativamente o produzca efectos jurídicos en el interesado.

La norma exige como requisito adicional, cuando se aplica alguna de las excepciones que establece el precepto en su apartado 2, salvo en caso de habilitación normativa, la adopción de medidas adecuadas para salvaguardar los derechos y libertades y los derechos e intereses legítimos del interesado: derecho a obtener intervención humana por parte del responsable, a expresar su punto de vista y a impugnar la decisión.

Por tanto, esta norma establece lo que se ha denominado una reserva de humanidad para determinadas actuaciones que puede ser excepcionado en

diversos casos, entre los que se encuentra que una norma específicamente prevea el uso de la IA, que debería tener rango de ley, por aplicación del artículo 18.4 CE. En nuestro ordenamiento jurídico dicha habilitación normativa podría entenderse comprendida en el artículo 41 de la LRJSP.

Ahora bien, el TJUE ha aclarado el alcance del concepto de «decisión basada únicamente en tratamiento automatizado», mediante la sentencia del caso C-634/2021, OQ contra Land Hessen, de 7 de diciembre de 2023, en términos no totalmente coincidentes con la definición que contiene el artículo 41 de la LRJSP. El TJUE con motivo de examinar un sistema privado de credit scoring (puntuaje crediticio) cuyos resultados eran utilizados por bancos para conceder o denegar créditos, declaró que la generación automatizada de un valor de solvencia -capacidad de una persona para hacer frente a futuros compromisos de pago-(score) por una agencia de crédito constituye una decisión individual automatizada en el sentido del artículo 22 del RGPD cuando dicho score determina de manera decisiva que un tercero celebre o rescinda un contrato con la persona. De manera que, aunque formalmente la decisión final la adopte un humano, si en la práctica el algoritmo tiene un peso determinante en esa decisión, la situación queda bajo el ámbito del artículo 22 del RGPD, es decir, la simple intervención nominal y formal de un humano no impide considerar la decisión automatizada a efectos jurídicos cuando se basa esencialmente en la recomendación automatizada.

Por tanto, debemos distinguir entre sistemas de mera verificación técnica automatizada y la toma de decisiones con impacto jurídico significativo. En los primeros la validación técnica de las condiciones preestablecidas como herramienta auxiliar en la toma de decisiones no implica decisiones automatizadas estrictamente prohibidas. Sin embargo, cuando la verificación automatizada se convierte en el único factor que determina una decisión que afecta a una persona procedería aplicar el régimen de protección reforzada del artículo 22 del RPD.

Además, el RGPD impone obligaciones de transparencia específicas respecto a los sistemas de decisión automatizada dirigidas a los afectados. Los artículos 13 y 14 disponen que si se prevé realizar decisiones automatizadas con esos datos se debe informar de tal hecho y proporcionar «información significativa sobre la lógica aplicada, así como la importancia y consecuencias previstas» de ese procesamiento automatizado. Asimismo, el artículo 15 establece el derecho de acceso del interesado a obtener esa información (véase el considerando 71).

En estos casos, el reglamento impone a los responsables del tratamiento que den “información significativa sobre la lógica aplicada, así como la importancia y las consecuencias previstas de dicho tratamiento para el interesado” (artículo 13.2. f) y artículo 14.2.g)) y reconoce a los afectados en derecho a acceder a

dicha información (artículo 15.1.h)), reproduciendo literalmente idéntica formulación que los preceptos antes referidos.

El TJUE ha afirmado que la finalidad principal del derecho previsto en el artículo 15 es permitir al afectado ejercer efectivamente los derechos que le reconoce el artículo 22, y en particular expresar su punto de vista e impugnar la decisión. Ello supone que el interesado deberá recibir una explicación de la decisión automatizada para poder recurrirla con el fin de que se encuentre garantizado su derecho de defensa, es decir, una explicación comprensible del procedimiento y los principios en que se basó la decisión automatizada que le afecta, como se deduce del considerando 71 del RGPD qué hace referencia al derecho a una explicación sobre el funcionamiento del mecanismo automatizado que produjo la decisión y sobre el resultado alcanzado.

En relación con ese derecho, el TJUE en la sentencia de 27 de febrero de 2025 (Asunto C-203/22, 1 Dun & Bradstreet Austria) hace hincapié en que dicha explicación debe presentarse de forma concisa, transparente e inteligible, e indicar que datos del interesado se utilizaron y cómo se emplearon en el proceso automatizado para llegar a ese resultado, sin que baste con revelar fórmulas matemáticas complejas ni con describir con extremo detalle cada fase del algoritmo, pues ello no sería ni conciso ni inteligible para el ciudadano medio. En esta sentencia se añade que si el responsable alegara que dar información sobre la lógica algorítmica revelaría secretos comerciales o datos de terceros, debía facilitar esa información confidencial a la autoridad de control (agencia de protección de datos) o al juez para que realizara una ponderación independiente de los intereses en juegos.

Las consideraciones de esta sentencia, acerca del alcance de la explicación, se encuentran íntimamente vinculadas con el deber de motivación de los actos administrativos, en este caso, fruto de un procedimiento administrativo de decisiones automatizadas. Cabe la posibilidad de que la motivación de la decisión provenga parcialmente de la configuración del algoritmo y su código fuente, por lo que en tal caso la administración debería ser capaz de traducir el razonamiento algorítmico a un lenguaje comprensible para el ciudadano con el fin de satisfacer su derecho a la explicación. Ahora bien, la sentencia se ocupa de clarificar que esa traducción no implica revelar todo el código fuente ni secretos técnicos detallados, sino explicar el criterio o criterios esenciales que se aplicaron por el sistema a los datos del individuo para alcanzar el resultado que se materializó en la decisión adoptada.

Naturalmente, esa explicación, que constituye también la motivación de la decisión adoptada, puede expresarse en la propia decisión automatizada, es decir, en la resolución administrativa, quedando así satisfecho tanto el derecho a la explicación reconocido por él RUEIA, como en derecho a la motivación de las resoluciones administrativas que reconoce nuestra legislación administrativa.

La interpretación integradora de sus artículos 15 y 22 del RGPD y de su considerando 71, a lo sumo, permite deducir la existencia de un derecho del afectado a que se le explique la concreta decisión automatizada de la que ha sido objeto. Destacamos al respecto que este derecho solo se reconocería a la concreta persona física objeto de la decisión y solo sería exigible frente a las decisiones totalmente automatizadas.

2.- Por otro lado, el **Reglamento (UE) 2024/1689, de 13 de junio de 2024, del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial** (**RIA)** tampoco recoge de forma expresa este derecho digital de transparencia sobre el uso por las Administraciones públicas de instrumentos de inteligencia artificial y sobre su funcionamiento.

En el ámbito de la Unión Europea, la preocupación por asegurar una inteligencia artificial (IA) fiable y centrada en el ser humano se ha concretado en el Reglamento (UE) 2024/1689, de 13 de junio de 2024 (RIA).

El Reglamento IA es la primera normativa integral en materia de IA en la UE y resulta pionera en el mundo.

El RIA establece un marco jurídico uniforme para la comercialización, despliegue y uso de sistemas de inteligencia artificial en la Unión Europea, adoptando un enfoque basado en el riesgo de los sistemas de IA, con el fin de garantizar que sean seguros, fiables y respetuosos con los derechos fundamentales e imponiendo requisitos a los proveedores (desarrolladores) y a los responsables del despliegue (usuarios finales de los sistemas, entre ellos las Administraciones Públicas).

El Reglamento establece normas armonizadas en materia de IA, imponiendo obligaciones específicas, en particular respecto de los sistemas de alto riesgo, con el fin de garantizar los derechos y libertades de las personas.

Su ámbito de aplicación se extiende a: proveedores de sistemas de IA (persona física o jurídica o entidad pública que desarrolla o para quien se desarrolla un sistema de IA y lo pone en servicio o lo comercializa bajo su nombre o marca, mediando un pago o no) que se pongan en servicio o comercialicen dentro de la UE o cuya salida se utilice en la UE, independientemente de su origen; y a responsables del despliegue o usuarios de los sistemas (persona física o jurídica, pública o privada, bajo cuya autoridad se utilice el sistema), considerando así usuarios a quienes explotan esos sistemas, y no a los afectados.

El RIA se basa en la IA generativa de propósito general y se aplica a cualquier tecnología creada o empleada en Europa.

El Reglamento de Inteligencia Artificial es una regulación basada en analizar el riesgo de los escenarios de uso, que se catalogan en cuatro niveles: (i) riesgo inaceptable; (ii) riesgo alto; (iii) riesgo limitado; y (iv) riesgo mínimo o nulo.

(i) En primer lugar, se prohíben las aplicaciones y los sistemas que crean un riesgo inaceptable, que están prohibidos, como la calificación social o el uso indiscriminado de identificación biométrica.
(ii) (En segundo lugar, se categorizan las aplicaciones de alto riesgo, las cuales deben estar sujetas a requisitos legales específicos (como, por ejemplo, una herramienta de escaneo del curriculum vitae que clasificara a los solicitantes de empleo para su aceptación o no).
(iii) (En tercer lugar, se detallan las aplicaciones que no están explícitamente prohibidas o catalogadas como de alto riesgo, las cuales quedan en gran parte sin regular. En el nivel 3 de riesgo limitado solo hay como requisito el comunicar que se interactúa con un sistema de IA (la denominada transparencia en el uso del sistema de IA).
(iv) En cuarto lugar, se contemplan los sistemas de riesgo mínimo o nulo, para los que no existen restricciones.

En este reglamento se describe el concepto sistemas de IA de alto riesgo (articulo 6) y se concretan en el anexo III, como aquellos sistemas de IA que se utilizan en cualquiera de los ocho siguientes escenarios de alto riesgo cuando la salida que producen sea relevante en una decisión con posible riesgo sobre la salud, la seguridad o los derechos fundamentales:

- Identificación biométrica y categorización de personas físicas;

- Gestión y funcionamiento de infraestructuras esenciales (tráfico, electricidad, agua, etc.);

- Educación y formación profesional (gestión del acceso, planificación académica);

- Empleo, gestión de los trabajadores y acceso al autoempleo;

- Acceso y disfrute de servicios públicos y privados esenciales y sus beneficios;

- Actividades de fuerzas y cuerpos de seguridad (como valoración de pruebas o de sospechosos);

- Gestión de la migración, el asilo y el control fronterizo; o

- Administración de justicia y procesos democráticos.

La Comisión puede actualizar esta lista mediante un acto delegado.

En estos escenarios, es necesario desarrollar sistemas de IA responsable que sean auditables, garanticen la seguridad de su comportamiento a lo largo del

ciclo de vida y tengan definidos procedimientos para minimizar el riesgo y previstos procedimientos que garanticen una reparación adecuada.

Concretamente, los proveedores de sistemas de alto riesgo deben cumplir las siguientes exigencias:

- Contar con un sistema de gestión de riesgos, que contemple, en particular, los riesgos sobre la salud, seguridad y derechos fundamentales relacionados con su propósito.

- Establecer una gobernanza y gestión de los datos de entrenamiento y prueba (asegurando buenas prácticas en su diseño, evitando sesgos que afecten negativamente a las personas, etc.).

- Los sistemas irán acompañados de documentación técnica actualizada, que demuestre que se cumplen los requisitos exigidos. Se especifica un contenido mínimo, que la Comisión puede enmendar. Y se recoge en el anexo iv, que podemos resumir del siguiente modo:

La descripción detallada de los elementos del sistema de IA y de su proceso de desarrollo debe incluir:

- los métodos y las medidas adoptados para el desarrollo del sistema de IA;

- las especificaciones de diseño del sistema, a saber, la lógica general del sistema de IA y de los algoritmos empleados;

- la arquitectura del sistema que detalle cómo se incorporan o se enriquecen mutuamente los componentes del software, cuando proceda;

- los requisitos sobre datos en forma de fichas técnicas que describan las metodologías y las técnicas de aprendizaje automático, así como los conjuntos de datos de entrenamiento utilizados, incluida la información acerca de la procedencia de dichos conjuntos de datos, su alcance y sus características principales;

- la evaluación de las medidas de vigilancia humana necesarias;

- en su caso, los cambios predeterminados en el sistema de IA y en su funcionamiento;

- los procedimientos de validación y prueba utilizados, incluida la información acerca de los datos de validación y prueba empleados y sus características principales.

- Los sistemas tomarán automáticamente registros de actividad del sistema.

- Se aportará información a los usuarios sobre las capacidades del sistema, sus requisitos de equipamiento, su ámbito de aplicación, su nivel de precisión, las condiciones de utilización que pueden implicar riesgos, los sistemas para supervisión humana, etc.

- Los sistemas permitirán la supervisión por personas durante su uso para minimizar los riesgos a la salud, seguridad y derechos fundamentales, en particular de los riesgos residuales tras la aplicación de medidas de mitigación. Los usuarios podrán monitorizar los sistemas e interpretar sus salidas. Para identificación biométrica remota, la salida requerirá verificación por una persona física, posiblemente dos.

- Los sistemas proporcionaran un nivel adecuado de precisión, robustez y ciberseguridad, que se declarará en la documentación que los acompaña.

De esta manera, el RIA se busca anticiparse a los problemas y supuestos de mayor riesgo, pretendiendo que los algoritmos sean transparentes, éticos, imparciales, explicables y estén bajo control humano.

En definitiva, procura la implantación de sistemas de IA responsable bajo el paradigma de la IA fiable auditable y segura.

Estos escenarios de riesgo requerirán normas o estándares para asegurar la calidad y la seguridad de los sistemas de IA, diseño de guías de uso de la IA que permitan a las empresas adherirse a la regulación y cumplirla y mecanismos de certificación. Todo ello permitirá avanzar hacia el diseño de metodologías para auditar los sistemas de IA y garantizar los principios de responsabilidad que deben regir el uso y el desarrollo de estos sistemas.

Al margen de esta clasificación de sistemas de riesgo, hemos de hacer referencia a una categoría adicional, los sistemas de IA de uso general, y definidos como aquellos entrenados con gran cantidad de datos, utilizando auto-supervisión a gran escala que son capaces de realizar de manera competente gran variedad de tareas distintas. Estos sistemas pueden ser utilizados como sistemas de IA de alto riesgo por sí solos o ser componentes de sistemas de IA de alto riesgo (considerando 85) y debe cumplir una serie de requisitos y obligaciones específicas, previstas en los artículos 51 a 56 del Reglamento.

Dicen así los Considerandos del Reglamento:

«(97) El concepto de modelos de IA de uso general debe definirse claramente y diferenciarse del concepto de sistemas de IA con el fin de garantizar la seguridad jurídica. La definición debe basarse en las características funcionales esenciales de un modelo de IA de uso general, en particular la generalidad y la capacidad de realizar de manera competente una amplia variedad de tareas diferenciadas. Estos modelos suelen entrenarse usando grandes volúmenes de datos y a través de diversos métodos, como el aprendizaje autosupervisado, no supervisado o por refuerzo. Los modelos de IA de uso general pueden introducirse en el mercado de diversas maneras, por ejemplo, a través de bibliotecas, interfaces de programación de aplicaciones (API), como descarga directa o como copia física. Estos modelos pueden modificarse o perfeccionarse y transformarse en nuevos modelos. Aunque los modelos de IA son componentes esenciales de los sistemas de IA, no constituyen por sí mismos sistemas de IA. Los modelos de IA requieren que se les añadan otros componentes, como, por ejemplo, una interfaz de usuario, para convertirse en sistemas de IA. Los modelos de IA suelen estar integrados en los sistemas de IA y formar parte de dichos sistemas. El presente Reglamento establece normas específicas para los modelos de IA de uso general y para los modelos de IA de uso general que entrañan riesgos sistémicos, que deben aplicarse también cuando estos modelos estén integrados en un sistema de IA o formen parte de un sistema de IA. Debe entenderse que las obligaciones de los proveedores de modelos de IA de uso general deben aplicarse una vez que los modelos de IA de uso general se introduzcan en el mercado. Cuando el proveedor de un modelo de IA de uso general integre un modelo propio en un sistema de IA propio que se comercialice o ponga en servicio, se debe considerar que dicho modelo se ha introducido en el mercado y, por tanto, se deben seguir aplicando las obligaciones establecidas en el presente Reglamento en relación con los modelos, además de las establecidas en relación con los sistemas de IA. En cualquier caso, las obligaciones establecidas en relación con los modelos no deben aplicarse cuando un modelo propio se utilice en procesos puramente internos que no sean esenciales para suministrar un producto o un servicio a un tercero y los derechos de las personas físicas no se vean afectados. Teniendo en cuenta su potencial para causar efectos negativos importantes, los modelos de IA de uso general con riesgo sistémico deben estar siempre sujetos a las obligaciones pertinentes establecidas en el presente Reglamento. La definición no debe incluir los modelos de IA utilizados antes de su introducción en el mercado únicamente para actividades de investigación, desarrollo y creación de prototipos. Lo anterior se entiende sin perjuicio de la obligación de cumplir lo

dispuesto en el presente Reglamento cuando, tras haber realizado dichas actividades, el modelo se introduzca en el mercado.»

«(100) Cuando un modelo de IA de uso general esté integrado en un sistema de IA o forme parte de él, este sistema debe considerarse un sistema de IA de uso general cuando, debido a esta integración, el sistema tenga la capacidad de servir a diversos fines. Un sistema de IA de uso general puede utilizarse directamente e integrarse en otros sistemas de IA.»

El término transparencia se emplea 41 veces en el texto del Reglamento, aunque no sea definido normativamente, si bien el RIA enuncia la transparencia como un principio fundamental de la regulación (art. 2 d)). El considerando 27 alude al significado que elaboró el grupo independiente de expertos de alto nivel sobre IA con motivo de la elaboración de las directrices éticas para una IA fiable en 2019, del siguiente modo: *«Por «transparencia» se entiende que los sistemas de IA se desarrollan y utilizan de un modo que permita una trazabilidad y explicabilidad adecuadas, y que, al mismo tiempo, haga que las personas sean conscientes de que se comunican o interactúan con un sistema de IA e informe debidamente a los responsables del despliegue acerca de las capacidades y limitaciones de dicho sistema de IA y a las personas afectadas acerca de sus derechos».*

Por lo que respecta a las obligaciones de transparencia, el artículo 50 establece obligaciones de transparencia para proveedores y responsables del despliegue, entendiéndose por estos, según el artículo 1: *«responsable del despliegue»: una persona física o jurídica, o autoridad pública, órgano u organismo que utilice un sistema de IA bajo su propia autoridad, salvo cuando su uso se enmarque en una actividad personal de carácter no profesional».*

De modo que los proveedores se asegurarán de que las personas físicas sean informadas de que están interactuando con un sistema IA, salvo que sea obvio, con ciertas excepciones para la persecución del crimen.

Los usuarios de sistemas o responsables del despliegue de categorización biométrica o de reconocimiento de emociones, informarán del funcionamiento del sistema a las personas sobre los que se use de tal realidad, también con ciertas excepciones para la persecución del crimen.

Asimismo, los responsables del despliegue de sistemas de IA que produzcan imagen o sonido que parezcan de verdad personas, lugares, objetos, etc. (deep fakes), deberán informar de ello, con ciertas excepciones en la persecución del crimen o en contenidos evidentemente creativos, satíricos o ficticios.

Así, el RIA dedica el capítulo IV a determinar una serie de obligaciones de transparencia cuyo beneficiario son las personas físicas a las que el sistema de inteligencia artificial puede afectar. Además, la información debe ser facilitada a las personas físicas de manera clara y distinguible; a más tardar con ocasión de la primera interacción o exposición con el sistema (art. 50.5 RIA).

Por otro lado, la clasificación de sistemas de IA en función de su riesgo conduce a distinguir las exigencias de transparencia en función de qué se trate de sistemas de riesgo medio, donde las previsiones de transparencia prácticamente

se circunscriben a advertir al usuario de la utilización de IA en la generación de contenidos, y sistemas de IA de alto riesgo dónde se somete a un mayor número de obligaciones de información y transparencia a los proveedores y responsables del despliegue.

En particular, en relación con estos sistemas de alto riesgo se exige que se diseñen y desarrollen de tal modo que quede garantizado que «funcionan con un nivel de transparencia suficiente para que los responsables del despliegue interpreten y usen correctamente su información de salida» (artículo 13.1); entendiendo por responsable del despliegue la *«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»* (artículo 3.4). Esta exigencia de transparencia se concreta en la obligación de que estos sistemas vayan acompañados de unas «instrucciones de uso» (artículo 13.2) en las que debe constar un largo listado de información relacionada con el funcionamiento y garantías del correspondiente sistema de IA de alto riesgo (artículo 13.3).

Por tanto, conforme se establece en el artículo 13.1 del Reglamento de IA, los sistemas de IA de alto riesgo deben ser diseñados y desarrollados de modo que permitan a los responsables del despliegue -personas que emplean el sistema bajo su propia autoridad para un uso profesional- comprender y evaluar el funcionamiento del sistema de IA, interpretar el resultado y utilizarlos adecuadamente.

No obstante, debe ponerse de manifiesto que las obligaciones de suministro de información que afectan a los sistemas de IA de alto riesgo encuentran un importante límite en la confidencialidad de la información. Así se deduce con claridad del artículo 78 del Reglamento que impone a las autoridades de vigilancia, a los organismos notificados y a la comisión respetar los derechos de propiedad intelectual e industrial, del secreto profesional o comercial y del secreto oficial.

A su vez el artículo 21.3 somete a las exigencias de confidencialidad toda información obtenida por una autoridad competente y el artículo 25.5 se refiere a los proveedores que cooperen por otros proveedores posteriores, a los que faciliten sus sistemas de IA, obligándoles a respetar los derechos de propiedad intelectual e industrial, la información empresarial confidencial y los secretos comerciales en el cumplimiento de su deber de proporcionar información relativa al uso de tales sistemas.

Sin duda, el Reglamento de IA contribuye al cumplimiento de las previsiones sobre transparencia establecidas el RGPD, la LOPD y la LTAIBG, combatiendo de este modo la opacidad en el uso de los sistemas de IA, en general, y en particular en la toma de decisiones automatizadas por parte de las Administraciones Públicas.

Así es, para el Reglamento la transparencia constituye un medio para que los usuarios y, en general, todos aquellos sujetos que intervienen en el ciclo de vida del producto en cualquier momento sean conocedores del funcionamiento interno de los sistemas de IA, así como para facilitar su supervisión por parte de las autoridades y organismos pertinentes, pero solo con el concreto alcance expresado en su texto que no incluye el acceso al algoritmo o código fuente de los sistemas de IA.

El Reglamento de IA, por su parte, establece también reserva de humanidad en tanto que prohíbe el uso de la IA en ciertos casos (artículo 5) y exige supervisión humana y obligaciones de transparencia para usuarios y personas afectadas en los sistemas de alto riesgo (artículo 14).

Por lo que respecta a la supervisión humana (artículo 14), el RIA establece la necesidad de control humano efectivo sobre los sistemas de alto riesgo, de manera que los proveedores deben diseñar la IA de forma que pueda ser vigilada y controlada por personas durante su uso, y los responsables del despliegue (usuarios finales, incluidas las autoridades públicas) tienen la obligación de establecer medidas de supervisión humana apropiadas antes de utilizar el sistema, lo que incluye garantizar que el personal encargado de la supervisión esté debidamente formado y facultado para intervenir, corregir o desactivar el sistema cuando sea necesario. De este modo lo que se pretende es que los usuarios de los sistemas comprendan sus recomendaciones y tengan capacidad para tomar decisiones informadas sobre si aceptar o modificar el resultado ofrecido por la decisión automatizada.

En relación con las exigencias de transparencia y explicaciones a usuarios -responsables del despliegue- y personas afectadas, el RIA impone que los sistemas de alto riesgo deban ir acompañados de instrucciones de uso claras para sus usuarios y de información que permite interpretar correctamente sus resultados.

En su artículo 86, titulado «derecho a explicación de decisiones tomadas individualmente», establece que toda persona afectada por una decisión de un responsable público, basada, principalmente, en la salida de un sistema de IA de alto riesgo que produzca efectos jurídicos o le afecte considerablemente, tiene derecho a obtener del responsable explicaciones claras y significativas sobre el papel que tuvo la IA en el proceso de decisión y sobre los principales elementos de la decisión final. Esta previsión, que aborda la explicabilidad de los sistemas de IA, es aplicable a los sistemas de IA de alto riesgo enumerados en el Anexo III (salvo los de categoría 2, que corresponden a sistemas de identificación biométrica remota en espacios públicos, que están prohibidos salvo excepciones concretas) y, por ende, a las decisiones automatizadas del sector público, e impone la obligación de explicar cómo influyó el algoritmo en el resultado. No obstante, el derecho de explicación del artículo 86 se aplica de forma subsidiaria

respecto de otras normas que puedan contemplar un derecho similar en la Unión Europea, reforzando el derecho de explicación en contextos de IA y complementando lo previsto en el RGPD.

Por tanto, si un ciudadano resulta perjudicado por una decisión automatizada de una Administración pública, tiene derecho a solicitar y recibir una explicación comprensible de la contribución o aportación de la IA a esa decisión, información que le permitirá ejercer sus derechos e impugnar, en su caso, la decisión.

Pero proporcionar a los afectados por sistemas de decisión automatizada "información significativa sobre la lógica aplicada, así como la importancia y consecuencias previstas" (artículos 13 y 14 del RGPD), el derecho de las personas afectadas por decisiones de responsables públicos basada en la salida de sistemas de IA a "explicación de decisiones tomadas individualmente" (artículo 86 RIA) o el cumplimiento de las obligaciones de transparencia que establece en favor de los responsables del despliegue el artículo 13 del RIA, no equivalen a exigir el acceso directo al algoritmo o a su código fuente -mediante una detallada explicación de los algoritmos utilizados o la revelación de todo el algoritmo-, básicamente, porque ese acceso no sería compatible con la tutela de los secretos comerciales y el derecho a la propiedad intelectual sobre los programas informáticos, cuya salvaguarda se exige expresamente en el considerando 63 del RIA, donde de forma taxativa se afirma que ese derecho de información no debe afectar negativamente a los derechos y libertades de terceros, incluidos los secretos comerciales o la propiedad intelectual que protege los programas informáticos.

De este modo, ambos reglamentos, RGPD y RIA, resultan complementarios en el establecimiento de obligaciones de transparencia y explicabilidad en relación con las decisiones automatizadas de las Administraciones públicas, cuyo cumplimiento resulta indispensable en el despliegue de los sistemas de IA, pero de ninguna manera implican un derecho de acceso al algoritmo o al código fuente de usuarios de los sistemas o de los afectados, y mucho menos, claro está, de los ciudadanos en general.

Tan solo se admite el acceso al código fuente del sistema de IA a las autoridades de vigilancia en el artículo 74.13 del RIA en los siguientes términos:

«13. Se concederá a las autoridades de vigilancia del mercado acceso al código fuente del sistema de IA de alto riesgo, previa solicitud motivada y solo si se cumplen las dos siguientes condiciones:

a) el acceso al código fuente es necesario para evaluar la conformidad de un sistema de IA de alto riesgo con los requisitos establecidos en el capítulo III, sección 2, y

b) se han agotado todos los procedimientos de prueba o auditoría y todas las comprobaciones basadas en los datos y la documentación facilitados por el proveedor, o han resultado insuficientes.

14. Cualesquiera información o documentación obtenidas por las autoridades de vigilancia del mercado se tratarán de conformidad con las obligaciones de confidencialidad establecidas en el artículo 78.»

Igualmente, en el artículo 92 se autoriza el acceso al código fuente a la Comisión para *«realizar evaluaciones del modelo de IA de uso general de que se trate con el fin de:*

a) evaluar si el proveedor cumple sus obligaciones en virtud del presente Reglamento, cuando la información recabada con arreglo al artículo 91 resulte insuficiente, o

b) investigar riesgos sistémicos a escala de la Unión de modelos de IA de uso general con riesgo sistémico, en particular a raíz de una alerta cualificada del grupo de expertos científicos de conformidad con el artículo 90, apartado 1, letra a)».

Por lo que respecta a la gobernanza de la IA, el Reglamento de Inteligencia Artificial establece la obligación de que cada Estado miembro designe al menos una autoridad de vigilancia del mercado, sin perjuicio de la existencia de autoridades específicas en determinados ámbitos. Por tanto, En Europa la gobernanza de la IA se lleva a cabo entre la UE y los Estados miembros, si bien el régimen sancionador se ha dejado en manos de los Estados, tal y como se establece el en artículo 99 del RIA, por el que se establecen normas armonizadas en materia de inteligencia artificial (Reglamento de Inteligencia Artificial) (se prevén multas que pueden llegar hasta el 7 % del volumen de negocios anual global de la empresa infractora o de 35 millones de euros).

La Agencia Española de Supervisión de Inteligencia Artificial (AESIA) es autoridad nacional encargada de supervisar la aplicación del RIA, coordinar las actuaciones de los Estados miembros, actuar como punto de contacto único con la Comisión Europea y representar al Estado español ante el Comité Europeo de Inteligencia Artificial.

En definitiva, el objetivo del Reglamento de Inteligencia Artificial, en lo que al presente concierne, es promover la adopción de una IA centrada en el ser humano y fiable, garantizando al mismo tiempo un elevado nivel de protección de la salud, la seguridad y los derechos fundamentales, estableciendo para ello requisitos específicos para los sistemas de IA de alto riesgo, obligaciones para los operadores de dichos sistemas y normas armonizadas de transparencia aplicables a determinados sistemas de IA.

El citado Reglamento de Inteligencia Artificial establece un estándar europeo de garantías –transparencia, supervisión humana efectiva, calidad y gobernanza– que resulta ineludible cuando se despliegan herramientas de IA en ámbitos directamente conectados con la tutela judicial efectiva y la imparcialidad.

En principio, el 2 de agosto de 2026 entrará en vigor Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (Reglamento de Inteligencia Artificial), que establece obligaciones específicas para sistemas de IA de alto riesgo empleados por autoridades públicas (artículo 113). Ahora

bien, determinadas exigencias para ciertos sistemas de alto riesgo no resultan exigibles hasta el 2 de diciembre de 2027 (artículo 6, apartado 1, y las obligaciones correspondientes del presente Reglamento).

3.- Por último, hemos de hacer una breve referencia **Convención Marco sobre IA del Consejo de Europa**, que establece obligaciones de transparencia sobre IA en el artículo 8.

Esta convención, firmada el 5 de septiembre de 2024, ya abierta a la firma de los estados, que entrará en vigor una vez hayan sido depositadas las ratificaciones de 5 estados, regula la inteligencia artificial desde el punto de vista de las finalidades básicas de la organización y, concretamente, la promoción y garantía de los derechos humanos, el Estado de Derecho y la democracia. Es decir, tiene por objetivo que el uso y actividades de la inteligencia artificial sea coherente con los derechos humanos, la democracia el Estado de Derecho.

Al igual que ocurre con el Reglamento IA es tecnológicamente neutral, porque no persigue regular una tecnología sino establecer garantías y obligaciones para que la actividad de la IA será respetuosa con los derechos humanos, la democracia el Estado de Derecho.

Define un sistema de IA como «un sistema basado en máquinas que, para objetivos explícitos o implícitos, deduce, de la información que recibe, cómo general resultados, como predicciones, contenidos, recomendaciones o decisiones que puedan influir en entornos físicos o virtuales. Los distintos sistemas de inteligencia artificial varían en sus niveles de autonomía y capacidad de adaptación tras su despliegue».

El Convenio es un instrumento complementario del Reglamento europeo de IA, si bien persiguen iguales objetivos y este último regula con mayor detalle cuestiones abordadas en el Convenio.

El Convenio aborda la cooperación intergubernamental sobre como regular el diseño, el desarrollo y la aplicación de sistemas de IA desde el punto de vista de su impacto en los derechos humanos, la democracia y el Estado de Derecho, mientras que el Reglamento es directamente aplicable en los estados miembros.

IV.4.- Una exigencia sustancial para las decisiones administrativas automatizadas: la motivación.

Debemos llamar la atención sobre la inexcusable necesidad de motivación de las decisiones administrativas automatizadas, en los términos que establece el artículo 35 de la LPAC.

Este deber de motivación también se impone por el artículo 86 del RIA. Además, si el sistema de IA solo se empleara como apoyo para la adopción de la decisión

administrativa, la motivación debería comprender el impacto que el uso del sistema algorítmico ha tenido en ella, expresando en qué medida se han aceptado las respuestas proporcionadas por la IA y por qué.

En efecto, en íntima conexión con el principio de transparencia surge el derecho de los ciudadanos a obtener de las Administraciones públicas a la motivación de las actuaciones administrativas, también cuando se trata de decisiones automatizadas, ex artículo 35 de la ley 39/2015, previsto como una de las garantías del procedimiento administrativo, que dimana de los artículos 9.3, 24, 103 y 106 de la Constitución.

Las exigencias del Estado de Derecho, conllevan que las decisiones administrativas automatizadas lleven aparejada la motivación adecuada de la decisión adoptada, expresando las razones que la justifican, de forma que resulte comprensible y congruente con el resultado, como impone la interdicción de la arbitrariedad (artículo 9.3 CE), el derecho de defensa (artículo 24 CE) y el deber de motivación administrativa (artículo 35 LPAC) y resulta de los principios de transparencia y buena administración.

De hecho, en la STS caso BOSCO la ausencia de motivación de la decisión ofrecida por el sistema automatizado tuvo trascendencia en la ponderación de intereses. Decíamos: «*la interpretación y ponderación de los límites del derecho de acceso a la información pública, particularmente, sobre una aplicación informática de toma de decisiones automatizadas, que esta Sala llevará a cabo a continuación, se encuentra condicionada al hecho de que ese funcionamiento automatizado sirve de soporte al reconocimiento o denegación de derechos sociales, arrojando un resultado positivo o negativo, sin exteriorizar las razones de dicho resultado».*

Ello no impide que también deban cumplirse las exigencias de explicabilidad y transparencia de los sistemas de IA en los términos exigidos por el RGPD, el RIA y la legislación española.

En particular, cuando se emplee como instrumento de apoyo un sistema de IA en la decisión administrativa, la motivación de la decisión deberá explicar el impacto qué el uso del sistema algorítmico ha tenido en esa decisión, ex artículo 86 del Reglamento de IA, y demostrar que no se han producido sesgos cognitivos.

Estas exigencias conllevan que no pueden ser utilizados algoritmos de caja negra en sistemas de IA no simbólica o estadística para automatizar la toma de decisiones administrativas sin supervisión humana, ante la imposibilidad de saber por qué el sistema adopta una u otra decisión.

Precisamente, cuando la complejidad de los sistemas de IA no simbólicos o estadísticos (aprendizaje automático o profundo, etc.) hace difícil explicar los pasos seguidos hasta la toma de la decisión final se impide el control judicial de la decisión adoptada y la interdicción de la arbitrariedad, ante la ausencia de debida motivación en la decisión administrativa adoptada.

Las consideraciones hasta aquí realizadas ponen de manifiesto qué, pese a que Reglamento de Protección de Datos y el Reglamento de inteligencia artificial establecen exigencias relativos al suministro de información acerca de los sistemas automatizados de toma de decisiones bajo determinadas condiciones y circunstancias, el derecho a la transparencia algorítmica con acceso al código fuente no encuentra respaldo en la normativa europea. De modo que la LTAIBG constituye el principal instrumento normativo para garantizar el acceso a los algoritmos y códigos fuentes de las aplicaciones y sistemas de inteligencia artificial empleados por las Administraciones Públicas en el ejercicio de sus potestades, ya sea para a adopción de decisiones totalmente automatizadas o como mero medio de apoyo en el proceso decisorio.

Como se ha puesto de manifiesto la LTAIBG no incluye disposiciones específicas sobre la transparencia de los algoritmos o de los sistemas de IA, sin perjuicio de que pueda accederse a esa información a través del ejercicio del derecho de acceso a la información pública, y con los límites que dicha disposición establece.

En definitiva, el control judicial sobre los sistemas de IA recae en los tribunales nacionales y en su caso en el Tribunal de Justicia de la Unión Europea (ante infracciones de derecho europeo) o el Tribunal Europeo de Derechos Humanos (ante infracciones del CEDH), por lo que si un ciudadano considera que una decisión automatizada pública ha violado sus derechos, puede acudir a la jurisdicción contenciosa administrativa correspondiente.

En tal caso, la autoridad judicial deberá requerir a la Administración la documentación del algoritmo y comprobar si la decisión vulneró el ordenamiento jurídico o, en su caso, si así se solicita podrá hacer efectivo el derecho de transparencia algorítmica con el alcance que se estime proporcional en cada caso, incluso, permitiendo el acceso al código fuente del algoritmo.

V. LOS LIMITES DEL ACCESO AL CODIGO FUENTE DE LOS ALGORITMOS COMO INFORMACIÓN PÚBLICA.

Antes de proseguir, recordemos que, tal y como declara nuestra jurisprudencia, siendo cierto que el derecho de acceso a la información pública no es ilimitado ni absoluto, al estar sometido a los límites contenidos en el artículo 14 y 15 de la LTAIBG, no lo es menos que tales límites deben aplicarse de forma justificada y proporcionada, ex artículo 14.2 de la LTAIBG, mediante una adecuada ponderación de los intereses en juego, el de acceso a la información pública, por un lado, y el protegido por la limitación de que se trate, por el otro, atendiendo a las concretas circunstancias del caso.

Asimismo, hemos de recordar, como hemos argumentado, que las causas de inadmisión y los límites al derecho de acceso contemplados en la LTAIBG deben interpretarse restrictivamente.

Dicho de otra forma, la concurrencia de un derecho subjetivo o interés legítimo para la invocación de un límite de los previstos en la LTAIBG no equivale mecánicamente a la denegación del acceso a la información pública, debiendo considerarse el perjuicio concreto al mismo y llevarse a cabo la debida ponderación de intereses en liza, bajo el presupuesto de que las limitaciones contempladas en la LTAIBG deben ser interpretadas de forma restrictiva y partiendo de la premisa de que el derecho de acceso a la información aparece configurado en nuestro ordenamiento con una formulación amplia, de manera que sólo son aceptables las limitaciones que resulten justificadas y proporcionadas.

A continuación, abordaremos algunos aspectos concretos de determinados límites al acceso a la información pública, de entre los expresados la LTAIPBG, tomando como referencia el debate surgido en torno a su alcance obstativo de tal acceso en el STS caso BOSCO.

La experiencia viene demostrando que los principales límites que se oponen a la transparencia algorítmica son la propiedad intelectual y la seguridad.

Ciertamente, también adquiere relevancia, como límite, la protección de datos de carácter personal, al que la LTAIPBG dedica el artículo 15, cuya regulación debe complementarse con lo previsto en el RGPD y la LPD.

La protección de datos personales de las personas físicas se configura como un derecho fundamental en el artículo 8.1 de la Carta de los Derechos Fundamentales de la Unión Europea, en el artículo 16.1 del Tratado de Funcionamiento de la Unión Europea y en el artículo 18.4 de la Constitución española.

El Reglamento General de Protección de Datos, complementado por la Ley Orgánica 3/2018, constituye, por tanto, el marco normativo aplicable en relación con la protección de los datos personales.

No obstante, lo natural es que el acceso al código fuente de los algoritmos de las aplicaciones empleadas por las Administraciones públicas, no implique el acceso a datos de carácter personal, sin perjuicio de que la potenciación de las vulnerabilidades de los sistemas pudiera implicar un posible acceso ilegítimo a tales datos.

Pues bien, en relación con la protección del derecho a los datos personales, debe precisarse que el límite previsto en el artículo 15 de la LTAIBG contempla específicamente la circunstancia de que la información solicitada por quien pretende ejercer el derecho de acceso a la información pública contuviera o

incluyera datos personales, ya se encuentren especialmente protegidos o no, lo que no parece que concurra cuando se pretende el acceso al código fuente de un algoritmo. Así se constató en el supuesto examinado en la STS caso BOSCO, donde el código fuente de la aplicación informática no contenía o incluía datos personales de los ciudadanos solicitantes de bono social, a quedar completamente al margen del acceso a la información pública, que pretendí la Fundación Ciudadana Civio.

V.1.- La propiedad intelectual (artículo 14.1.j) de la LTAIPBG).

Las aplicaciones telemáticas o electrónicas o los programas de ordenador son creaciones susceptibles de ser objeto de propiedad intelectual (artículos 10.1.i) y 95 y siguientes del Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia -TRLPI-). De modo que su código fuente queda específicamente protegido por el derecho de autor en la medida en la que constituye una forma de expresión del programa de ordenador, de conformidad con la jurisprudencia del Tribunal de Justicia de la Unión Europea, en interpretación del artículo 1 de la Directiva 2009/24/CE del Parlamento Europeo y del Consejo, de 23 de abril de 2009, sobre la protección jurídica de programas de ordenador, y sus precedentes (vid. SSTJUE de 22 de diciembre de 2010 (asunto C-393/09), apartados 34 y 35; de 6 de octubre de 2021 (asunto C-13/20), apartados 35 y 36; y de 17 de octubre de 2024 (C-159/23), apartados 37 y 38).

Sentado lo anterior, debemos distinguir aquellos supuestos donde los programas de ordenador empleados por las Administraciones públicas para la adopción de decisiones automatizadas han sido creados por estas de aquellos otros en que son de propiedad privada.

En este segundo supuesto, se ha cuestionado por la doctrina si cabe calificar estos programas de ordenador o aplicaciones telemáticas como una disposición general o un acto administrativo, con el objeto de aplicar el artículo 13 del Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual (TRLPI): *«No son objeto de propiedad intelectual las disposiciones legales o reglamentarias y sus correspondientes proyectos, las resoluciones de los órganos jurisdiccionales y los actos, acuerdos, deliberaciones y dictámenes de los organismos públicos, así como las traducciones oficiales de todos los textos anteriores».*

Así es, existe un debate en la doctrina sobre la naturaleza jurídica de los algoritmos creados por las administraciones públicas, pues algunos autores entienden que un sistema de IA o una aplicación informática que adopta decisiones totalmente automatizadas que emplea algoritmos que suponen la traducción a código informático de una norma previa, para su aplicación digital en el futuro en infinidad de ocasiones, debe ser considerada como disposición general, no mera aplicación instrumental o auxiliar. Y ello, aceptando que no

todos los algoritmos son siempre reglamentos y atribuyendo tal naturaleza solo a aquellos desarrollan una ley al traducirla a lenguaje formal para su lectura por máquinas (la STS caso Bosco consideró que la aplicación telemática empleada en ese caso era una herramienta tecnológica de carácter instrumental, que no cabía calificar ni como disposición general ni como acto administrativo).

En general, la protección jurídica que proporciona la propiedad intelectual y, particularmente, las facultades patrimoniales que integran el derecho de autor, viene justificada, en esencia, por la necesidad de defender y remunerar el trabajo y valor añadido que aporta el creador, así como la inversión de recursos de diversa naturaleza efectuada a tal efecto, otorgando un monopolio de disposición y explotación temporal que permita recuperar los costes incurridos e incentive su continuación como elemento fundamental del progreso cultural y técnico.

Sin embargo, dichas finalidades se presentan notoriamente atenuadas cuando el programa de ordenador ha sido creado por la propia Administración Pública, que es la titular de la propiedad intelectual, por mandato de la normativa del sector eléctrico para el ejercicio de competencias públicas y dirigida a servir a intereses igualmente públicos, no encontrándose, en consecuencia, integrada -o no, al menos, principalmente- en la lógica competitiva del mercado donde se proyectan con especial significación los derechos de explotación de la propiedad intelectual (en este sentido se pronunció la STS caso Bosco).

Además, en la hipótesis de que las aplicaciones telemáticas no fueran creadas por las Administraciones públicas, la protección de la propiedad intelectual como límite del acceso a los algoritmos podría proyectarse no tanto en la fase de acceso a la información sino en garantizar que no se haga una utilización posterior de la misma contraria a los intereses del autor de la obra y, en particular, a sus derechos de explotación, eventualidad ante la que se podrían adoptar determinadas cautelas que garantizarán la protección de ese derecho.

En particular, con el fin de minimizar los eventuales perjuicios que pudieran dimanar del acceso al código fuente como consecuencia de su explotación no autorizada por terceros, cabría someter el acceso a determinadas cautelas, como, por ejemplo, la prohibición de la difusión o la utilización del código fuente para otras finalidades sin la autorización expresa de la Administración, la advertencia expresa de la responsabilidad en que puede incurrir el solicitante de acceso por el incumplimiento de esa prohibición, la firma de un compromiso de uso limitado de la información recibida o la imposición de un deber de reserva o confidencialidad respecto de la información consultada. Este razonamiento fue empleado en la STS caso Bosco, al realizar la ponderación de intereses en juego, otorgando prevalencia al interés en el acceso al código fuente de la aplicación telemática BOSCO sobre el derecho a la propiedad intelectual de la Administración del Estado.

Las consideraciones anteriores se han dicho desde la perspectiva del uso por parte de las Administraciones Públicas de aplicaciones informáticas o sistemas de IA creados por ellas mismas. Pero lo cierto es que la mayoría de las Administraciones carecen de los medios técnicos y económicos suficientes para desarrollar por sí mismas sistemas de esta naturaleza. Por ello, habitualmente acuden a la contratación administrativa para obtener estas herramientas.

Debe reconocerse que en estos casos el límite de la propiedad intelectual puede jugar con una intensidad muy diferente a aquellos otros, en los que la aplicación informática o el sistema de IA forma parte del patrimonio de la Administración pública, dados los amplios términos en que el derecho a la propiedad intelectual se configura en el artículo 2 del texto refundido de la Ley de Propiedad intelectual que se encuentra integrada por derechos de carácter personal y patrimonial, que atribuyen al autor la plena disposición y el derecho exclusivo a la explotación de la obra, sin más limitaciones que las establecidas en la Ley. Se distinguen así dos bloques de derechos subjetivos integrados en el derecho a la propiedad intelectual que se desarrollan en el capítulo III del título II del libro primero del texto refundido, donde dedica la Sección Primera al «derecho moral» y la Sección Segunda a los «derechos de explotación», comprendiendo como contenido del derecho moral el de «decidir si su obra ha de ser divulgada y en qué forma» y entendiéndose por divulgación de la obra, tal y como establece el artículo 4 del texto refundido, «toda expresión de la misma que con el consentimiento del autor la haga accesible por primera vez al público en cualquier forma»

Parece, por tanto, que el derecho de divulgación, integrante del genérico derecho a la propiedad intelectual, comprende esencialmente la capacidad del autor de decidir si se permitirá el acceso al público a su obra por vez primera y, por ende, el acceso a la obra (algoritmo o código fuente), con independencia de su utilización o empleo posterior, exigiría el consentimiento del titular de la obra, entrando en juegos el límite del artículo 14.1.j) de la LTAIPBG.

En todo caso, una aceptando que el límite al acceso a la información pública del derecho a la propiedad intelectual opere con mayor intensidad cuándo la aplicación informática empleada por las Administraciones Públicas sea de titularidad privada, que en aquellos casos en los que estas crean sus propias herramientas informáticas, el procedimiento a seguir para la obtención de estas aplicaciones debe sujetarse, por lo general, a la contratación de servicios con empresas privadas que debería conllevar la cesión de los derechos de propiedad intelectual a la Administración, conforme se establece en el artículo 308.1 de la Ley de Contratos del Sector Público, cuya aplicabilidad viene determinada por el hecho de que «los contratos de adquisición de programas de ordenador desarrollados a medida» sean calificados por el artículo 16.b) de esa ley como contratos de servicios.

Además, en el marco del procedimiento de contratación regiría el artículo 133.1 de la Ley de Contratos del Sector Público, en virtud del cual en el procedimiento de adjudicación de los contratos *«los órganos de contratación no podrán divulgar la información facilitada por los empresarios que estos hayan designado como confidencial en el momento de presentar su oferta».*

Sin embargo, con independencia de lo expuesto, no es inusual que los contratistas impongan cláusulas de confidencialidad a las Administraciones contratantes en los contratos para la adquisición de sistemas de IA, supuesto este en el que el secreto comercial o empresarial o «los intereses económicos y comerciales» (artículo 14.1.h) LTAIPBG) se revelan como otro límite frecuentemente utilizado para obstaculizar el acceso al código fuente de las aplicaciones informáticas desarrolladas por empresas privadas y suministradas a las Administraciones Públicas, estrechamente vinculados a la libertad de empresa y al derecho a la propiedad intelectual consagrados por los artículos 16 y 17 dos de la carta de derechos fundamentales de la Unión Europea (STJUE, de 23 de noviembre de 2016, Bayer CropScience SA-NV y Stichting De Bijenstichting (asunto C-442/14, apartado 97 y ss).

La cuestión es encontrar un equilibrio entre la protección de los secretos comerciales y los derechos de propiedad intelectual de los contratistas y la transparencia en la actividad algorítmica administrativa. Tarea difícil dada la presión de las empresas tecnológicas con recursos y medios para desarrollar sistemas de IA para preservar el secreto comercial inherente al diseño del algoritmo y el código fuente, por razones competitivas.

En estos escenarios la explicabilidad de los sistemas de IA empleados por las Administraciones públicas y adquiridos mediante sistemas contratación pública, por lo general, no justificará el acceso al algoritmo o el código fuente, cumpliéndose tal exigencia con la divulgación de la propia existencia del sistema, los datos que utiliza, el proceso que sigue en su funcionamiento y las razones que motivan las decisiones que toma.

En todo caso, las Administraciones públicas también podrían promover procedimientos de licitación para el desarrollo de algoritmos de código abierto o desarrollar sus propios sistemas, aunque en estos casos se enfrentaría a la merma de seguridad de los sistemas de IA y al encarecimiento de estos sistemas, dado que el contratista se vería privado de la posibilidad de ofrecer el mismo sistema en el mercado.

En consecuencia, cuándo en la contratación pública de estas aplicaciones informáticas se introduzcan cláusulas de confidencialidad que excluyan el acceso al algoritmo o al código fuente, el derecho de acceso a la información pública se encontrará limitado por su virtualidad, sin perjuicio de la ponderación de intereses que deba hacerse en cada caso para evaluar la trascendencia de dicho límite.

V.2.- La seguridad pública (artículo 14.1.d) de la LTAIBG) y otros límites conexos.

Podemos aceptar que la revelación del código fuente de un algoritmo aumenta de una manera objetiva la severidad de las vulnerabilidades de cualquier aplicación informática, es decir, se trata de un riesgo inherente al acceso al código fuente, con carácter general.

Por ello, en la ponderación de derechos e intereses no pueden desdeñarse los riesgos de seguridad que pudiera generar el acceso de terceros al código fuente del algoritmo del sistema informático por las vulnerabilidades que entrañe. Pero tampoco puede soslayarse que estos riesgos, por lo general, pueden ser previstos, lo que posibilita el diseño de la aplicación o programa informático fortaleciendo la seguridad del sistema, con su consiguiente minimización.

Ciertamente, es probable que el acceso al código fuente erosione la eficacia del sistema automatizado de decisiones, obstaculizándose, su funcionamiento adecuado, al resultar posible para los afectados, una vez conocidos los indicadores o parámetros considerados por el sistema, sortearlos.

Este es un potencial riesgo que debe ser considerado pero aquella afirmación impide considerar que el riesgo de vulnerabilidad inherente al acceso al código fuente pueda oponerse, sin más, al derecho de acceso a la información pública que entraña el algoritmo. Entender otra cosa, vaciaría de contenido el derecho de acceso en relación con las aplicaciones telemáticas cuando tuviera por objeto el código fuente, e implicaría la atribución de un carácter absoluto a la limitación de "seguridad pública" que prevé el artículo 14.1.d) de la LTAIBG, sin duda, no previsto por el legislador.

A todo ello debemos añadir que, aun cuando en abstracto el acceso al código fuente de un programa pudiera incrementar potencialmente algunos riesgos sobre la seguridad informática de la aplicación, también cabe afirmar, en sentido contrario, que la transparencia sobre el mismo puede contribuir, en iguales términos potenciales, a la mejora del código y fortalecimiento de su seguridad puesto que, por un lado, incentiva a la Administración a extremar las cautelas de seguridad en el propio diseño y control del programa informático y, por otro lado, su escrutinio por actores diversos e independientes permite aflorar vulnerabilidades inicialmente inadvertidas y posibilitar su corrección temprana.

De hecho, en la actualidad no es insólita la disponibilidad de aplicaciones informáticas a través de licencias de código abierto y, particularmente, en el ámbito de aplicaciones desarrolladas por Administraciones Públicas para el ejercicio de sus competencias se encuentran ejemplos de publicación del código fuente, pudiendo citar, por su relevancia y notoriedad, las desarrolladas en el ámbito de la crisis sanitaria de la COVID-19 para el rastreo de personas infectadas y, en el concreto caso de España, la aplicación "Radar COVID" (disponible en https://github.com/RadarCOVID) sin que se entendiera que los

riesgos inherentes al conocimiento del código fuente o la naturaleza de la información concernida impidieran, de plano, su publicación.

Es más, aun cuando existen evidentes intereses y derechos relacionados con la confidencialidad, la protección de datos personales y la seguridad informática que deben quedar preferentemente tutelados cuando las circunstancias específicas de cada caso así lo aconsejen, es apreciable que, tanto en la normativa de la Unión Europea, como en la normativa doméstica existen mandatos y principios favorables a la transparencia de los algoritmos públicos que conducen a descartar la ocultación del código fuente como principio general y categórico de seguridad de los sistemas informáticos.

Es destacable, en este sentido, el Reglamento (UE) 2024/903 del Parlamento Europeo y del Consejo, de 13 de marzo de 2024 por el que se establecen medidas a fin de garantizar un alto nivel de interoperabilidad del sector público en toda la Unión (Reglamento sobre la Europa Interoperable) que resalta en sus considerandos 26 y 36 lo siguiente:

«*(26) Cuando las administraciones públicas compartan sus soluciones con otras administraciones públicas o con la ciudadanía, actúan en el interés público. Esto es aún más pertinente en el caso de las tecnologías innovadoras. Por ejemplo, el código abierto hace que los algoritmos sean transparentes y permite que se realicen auditorías independientes y se disponga de módulos reproducibles. El intercambio de soluciones de interoperabilidad entre administraciones públicas debe sentar las bases para crear un ecosistema abierto de tecnologías digitales para el sector público que pueda reportar múltiples beneficios.*

[...]

(36) Dado que el código abierto permite a los usuarios evaluar e inspeccionar activamente la interoperabilidad y seguridad de las soluciones, es importante que sustente la implantación de soluciones de interoperabilidad. En este contexto, y para mejorar la claridad jurídica y el reconocimiento mutuo de licencias en los Estados miembros, debe fomentarse el uso de licencias de código abierto. Con la Licencia Pública de la Unión Europea (EUPL, por sus siglas en inglés) la Comisión ya ofrece una solución para dicho tipo de concesión de licencias. Los portales de los Estados miembros que recojan soluciones de código abierto vinculadas al portal de la Europa Interoperable deben permitir el uso de la EUPL, aunque no se excluye la posibilidad de que dichos portales puedan permitir el uso de otras licencias de código abierto».

Y en su artículo 5, relativo a los principios generales, dispone:

«1. La Comisión publicará las soluciones de la Europa Interoperable y el Marco Europeo de Interoperabilidad en el portal de la Europa Interoperable por medios electrónicos, en formatos abiertos, legibles por máquina, accesibles a personas con discapacidad con arreglo a las Directivas (UE) 2016/2102 (19) y (UE) 2019/882 (20) del Parlamento Europeo y del Consejo, fáciles de encontrar y reutilizables, en su caso, junto con su código fuente documentado y sus metadatos. Las versiones de las soluciones de la Europa Interoperable

traducidas automáticamente se publicarán en el portal de la Europa Interoperable en todas las lenguas oficiales de las instituciones de la Unión».

En la misma materia de reutilización, el artículo 157.2 de la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público contempla la oportunidad de declarar determinadas aplicaciones de las Administraciones como de fuentes abiertas para aumentar la transparencia en su funcionamiento o fomentar la incorporación de los ciudadanos a la Sociedad de la información.

Por su lado, la normativa de protección de datos impone determinadas exigencias de información y transparencia sobre la lógica empleada cuando se realicen tratamientos de datos personales que impliquen la adopción de decisiones individuales automatizadas, esto es, decisiones basadas únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzcan efectos jurídicos en el interesado o le afecte significativamente de modo similar.

En estos casos, el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos, y por el que se deroga la Directiva 95/46/CE (RGPD), contempla el derecho del titular de la información personal a obtener, tanto al recabarse los datos (artículo 13.2.f y 14.2.g), como cuando se ejercita el derecho de acceso (artículo 15.1.h) « *información significativa sobre la lógica aplicada, así como la importancia y las consecuencias previstas de dicho tratamiento para el interesado.*».

Cabe precisar que esta exigencia del RGPD de "explicabilidad" de la lógica de las decisiones automatizadas no puede identificarse con la obligación de dar acceso al código fuente, pero tampoco queda en todo caso excluido ni se contrapone necesariamente con la seguridad en el tratamiento de los datos personales, como lo evidencian las recomendaciones adoptadas por el Comité Europeo de Protección de Datos en sus Directrices 04/2020 sobre el uso de datos de localización y herramientas de rastreo de contactos en el contexto de la pandemia de COVID-19, adoptadas el 21 de abril de 2020, en las que se afirmaba que:

«37. Para asegurar su equidad, la rendición de cuentas y, más en general, su consonancia con la ley, los algoritmos deben ser auditables y han ser revisados periódicamente por expertos independientes. El código fuente de la aplicación debe hacerse público con miras a un control lo más amplio posible».

«*GEN-3 El código fuente de la aplicación y de su servidor final debe ser abierto, y las especificaciones técnicas han de hacerse públicas, de modo que cualquier parte interesada pueda auditar el código y, cuando proceda, contribuir a mejorarlo, corrigiendo posibles errores y asegurando la transparencia en el tratamiento de datos personales.*».

En el mismo sentido, cabe citar el artículo 23 de la Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación que, en orden a tutelar el derecho de igualdad, contiene también principios favorables a la transparencia

de los algoritmos empleados en la toma de decisiones por las Administraciones Públicas en los siguientes términos:

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

2. Las administraciones públicas, en el marco de sus competencias en el ámbito de los algoritmos involucrados en procesos de toma de decisiones, priorizarán la transparencia en el diseño y la implementación y la capacidad de interpretación de las decisiones adoptadas por los mismos. [...]».

Lo dicho hasta el momento evidencia que la sola invocación de que la revelación del código fuente aumenta de una manera objetiva la severidad de las vulnerabilidades de cualquier aplicación informática es, en sí misma, insuficiente para excluir el acceso al mismo. Al margen de que dicho acceso puede también contribuir, en sentido opuesto, al robustecimiento de la seguridad, existen otros intereses de alta significación jurídica relacionados con la participación en la toma de decisiones, el fortalecimiento de la democracia, la efectividad de otros derechos constitucionales, la generación de confianza en las instituciones públicas y el aumento de la eficiencia y eficacia de la actuación pública que deben también tutelarse y tomarse en justa consideración en la ponderación que exige la Ley.

Por lo demás, la apreciación del mencionado riesgo de vulnerabilidad informática como obstáculo al acceso al código fuente, con carácter general, resultaría contrario a la propia exigencia de juicio de proporcionalidad y ponderación de intereses en juego que la LTAIBG exige. Así es, el límite de seguridad pública solo puede impedir el acceso a la información pública cuando ello se encuentre justificado y resulte proporcionado a su objeto y finalidad, previa ponderación de los intereses en conflicto, ex artículo 14.2 de la LTAIBG.

Por tanto, el debate sobre la aplicación de este límite debe centrarse en su ponderación y la de otros límites relacionados con el mismo, como el de la **«prevención, investigación y sanción de ilícitos penales, administrativos o disciplinarios»** (artículo 14.1.e) de la LTAIPBG) que, como consecuencia de las vulnerabilidades que genera el acceso al código fuente de la aplicación, pudieran verse comprometidos.

Resulta ilustrativo el caso analizado por la Resolución CTBG 810/2023, de 2 de octubre, en el qué aunque no se pretendía acceder al algoritmo se quería conocer información estrecha y directamente relacionada con el funcionamiento

del sistema Viogen, concretamente su cuestionario y el protocolo para su uso, acceso a la información que había sido denegada por la Administración con sustento en el límite del artículo 14.1.e) de la LTAIPBG dado el riesgo que entrañaba que se conocieran los parámetros empleados, la identificación de los riesgos y su análisis de vulneración de los sistemas de protección de las víctimas de violencia de género, siendo ratificada la decisión administrativo por el Consejo apelando al principio de cautela y prevalencia de los bienes jurídicos afectados frente al derecho de acceso.

Sin embargo, en la Resolución CTBG 910/2023 se autorizó el acceso a las funcionalidades y especificaciones técnicas de la aplicación VeriPol, porque no se había justificado suficientemente la aplicabilidad del límite.

En relación con la seguridad adquiere también trascendencia el límite de las **«funciones administrativas de vigilancia, inspección y control»** (artículo 14.1.g) de la LTAIPBG), ante los efectos que pudiera tener el acceso a la información pública de las aplicaciones informáticas sobre ámbitos de inspección, como la laboral o la tributaria. Al respecto en la Resolución CTBG 825/2019, de 13 de febrero, concedió el acceso al contenido y funciones de determinadas aplicaciones informáticas empleadas por la AEAT ZÚJAR, TESEO, INEX, INTER, DEDALO, PROMETEO y GENIO, entendiendo que la respuesta ofrecida por la Administración había sido genérica e insuficiente y que no había argumentado mínimamente la aplicación de los límites invocados para negar el acceso, en concreto, e) La prevención, investigación y sanción de los ilícitos penales administrativos o disciplinarios. g) Las funciones administrativas de vigilancia, inspección y control y j) El secreto profesional y la propiedad intelectual e industrial.

Asimismo, en la Resolución CTBG 46/2024 se concedió el acceso a información sobre funcionalidades y especificaciones técnicas de la aplicación MAX empleada para la supervisión de horas extraordinarias en el ámbito laboral, porque la Administración no había justificado suficientemente la concurrencia de límite invocado genéricamente.

Por último, frente al acceso a los algoritmos cabría invocar el límite previsto en el artículo 14.1.k) de la LTAIPBG, relativo a «**La garantía de la confidencialidad o el secreto requerido en procesos de toma de decisiones».** Este límite, cuya concurrencia exige que la información pretendida permita el acceso a datos que pudiera perjudicar de forma razonable y no meramente hipotética el proceso de toma de decisiones y su conclusión (Resolución CTBG 10/2019), ha sido invocado en algunas ocasiones, concretamente en relación con las aplicaciones PROGRES PRESS-CARPA-ALFA, empleada para el cálculo de la pensión, e INCA, utilizada para dar soporte a la gestión de prestaciones de incapacidad temporal, nacimiento y cuidado de menor, riesgo ante el embarazo y otras prestaciones, pero en los supuestos señalados la Administración no había

justificado suficientemente la aplicabilidad del límite invocado (Resoluciones CTBG 907/2023 y 908/2023).

En relación con este último límite, parece que no sería de aplicación a los supuestos de decisiones totalmente automatizadas cuándo son el resultado el ejercicio de potestades regladas de las Administraciones Públicas, dónde el algoritmo ha de reflejar la traducción a lenguaje informático de las normas jurídicas aplicables en cada caso, limitándose a comprobar el cumplimiento de los requisitos exigidos por la normativa aplicable. En estos casos, dado el carácter reglado de las decisiones y la ausencia de discrecionalidad alguna de la Administración, no cabría la posibilidad de que el acceso al algoritmo pudiera perjudicar en modo alguno el proceso decisorio o su conclusión.

Tampoco parece sencillo justificar su aplicación a aplicaciones informáticas empleadas como mecanismos de apoyo en la toma de decisiones administrativas pues no resulta fácil advertir en qué circunstancias el acceso a la información algorítmica podría perturbar la confidencialidad por el secreto en la toma de decisiones, si bien no resulta descartable, a priori.

VI.- LA TRANSPARENCIA ALGORÍTMICA Y EL CASO BOSCO

Volvamos ahora a la **STS num. 1119/2025, de 11 de septiembre de 2025 (rec. 7878/2024),** denominada caso BOSCO, para abordar el examen de un supuesto concreto de transparencia algorítmica con acceso al código fuente.

En la actualidad, hay **otro caso** pendiente de resolver en el Tribunal Supremo, relativo a la denegación del acceso al código fuente de una aplicación de sorteo de tribunales de oposiciones de la Comunidad de Madrid (Auto de la Sección Primera de la Sala Tercera del Tribunal Supremo de 10 de septiembre de 2025, recurso de casación núm. 3998/2025, donde el recurrente es el Consejo de Transparencia y Buen Gobierno).

A/ Sentencia recurrida

Sentencia de 30 de abril de 2024, dictada por la Sección Séptima de la Sala de lo Contencioso-Administrativo de la Audiencia Nacional, que desestima el recurso de apelación n.º 51/2022, presentado frente a la Sentencia de 30 de diciembre de 2021, dictada por el Juzgado Central de lo Contencioso-Administrativo n.º 8 en el procedimiento ordinario número 18/2019, por la que se desestimó el recurso contencioso-administrativo interpuesto por la Fundación Ciudadana Civio contra la resolución del Consejo de Transparencia y Buen Gobierno (CTBG) de 18 de febrero de 2019, dictada en el expediente R/0701/2018, por la que se estimó parcialmente la reclamación presentada contra el Ministerio de Transición Ecológica (MITECO) sobre acceso a la

información relativa a la aplicación telemática (BOSCO), desarrollada para determinar el cumplimiento de los requisitos para ser beneficiario del bono social.

B/ Antecedentes

1.- Fundación Ciudadana Civio solicitó, a través de Portal de Transparencia, la siguiente información sobre la aplicación telemática "BOSCO".

- La especificación técnica de dicha aplicación.

- El resultado de las pruebas realizadas para comprobar que la aplicación implementada cumple la especificación funcional.

- El código fuente de la aplicación actualmente en producción.

- Cualquier otro entregable que permita conocer el funcionamiento de la aplicación.

La aplicación BOSCO es la herramienta informática -plataforma informática- disponible en la sede electrónica del ministerio que deben emplear las empresas comercializadoras de energía eléctrica para conocer que consumidores, de entre los solicitantes del bono social, cumplen los requisitos, legal y reglamentariamente previstos, para ser considerados consumidores vulnerables (con derecho a gozar de un descuento en la tarifa eléctrica por razones de renta o circunstancias especiales - discapacidad, víctima de violencia de género, víctima de terrorismo, situación de dependencia, único progenitor que se hace cargo del menor y situación de electrodependencia- o pertenencia a determinados colectivos familias numerosas, pensionistas del Sistema de la Seguridad Social por jubilación o incapacidad permanente que perciban la cuantía mínima de pensión y beneficiarios del Ingreso Mínimo Vital-) o consumidores vulnerables severos (rentas más bajas y consumidores en riesgo de exclusión social -atendidos por servicios sociales-) y percibir, en consecuencia, el bono social, quedando sujetas a ese resultado (se limitan a introducir en la aplicación telemática los datos correspondientes al consumidor y la aplicación permite la visualización del resultado de las comprobaciones realizadas por la misma, determinándose de forma automatizada si un determinado consumidor tiene derecho al bono social). Si el consumidor concernido no está de acuerdo con el resultado de la operación telemática puede reclamar ante los servicios de consumo correspondientes, en los términos que establece la normativa de defensa de los consumidores, tal y como dispone el artículo 8 del Real Decreto 897/2017, de 6 de octubre

2.- Entendiéndose desestimada por silencio administrativo la solicitud, la Fundación Ciudadana Civio presentó reclamación ante el Consejo de Transparencia y Buen Gobierno (CTBG), que mediante Resolución de fecha 18 de febrero de 2019 estimó en parte la reclamación, estima parcialmente la reclamación autorizando el acceso a la especificación técnica de dicha aplicación, al resultado de las pruebas realizadas para comprobar que la aplicación implementada cumple la especificación funcional y cualquier otro entregable que permita conocer el funcionamiento de la aplicación, pero no concede el acceso al código fuente.

El CTBG analizó los límites contemplados en el artículo 14.1, letras a) y d) (seguridad nacional y seguridad pública) y rechazó que resultaran aplicables en el presente caso. Por el contrario, sí consideró que el código fuente estaba amparado por el derecho de propiedad intelectual (límite del artículo 14.1.j de la LTAIBG).

3.- Presentado recurso contencioso-administrativo por la Fundación Ciudadana Civio, fue desestimado por el JCCA y el recurso de apelación fue también desestimado por la Sala CA de la AN, que, en síntesis, considera improcedente facilitar el código fuente en aplicación del 14.1 j) de la LTAIBG, relativo a perjuicios a la propiedad intelectual, reconociendo el derecho de la Administración a la propiedad intelectual del programa de ordenador, a lo que añade que la entrega del código fuente de la aplicación BOSCO pondría en grave riesgo derechos de terceros y atentaría a bienes jurídicos protegidos por los límites al derecho de acceso a la información pública del artículo 14.1 letras d) seguridad pública, g) funciones administrativas de vigilancia, inspección y control, i) política económica y monetaria y k) garantía de confidencialidad o secreto en procesos de toma de decisión, de la LTAIBG (se podrían producir vulnerabilidades para acceder a las bases de datos conectadas con la aplicación que contienen datos especialmente protegidos, con sustento en el informe del Subdirector General de Tecnologías de la Información y las Comunicaciones del Ministerio de Industria, Comercio y Turismo).

C/ Criterio de la sala sobre la ponderación de los intereses en juego: los límites al acceso a la información pública.

Por otro lado, la interpretación y ponderación de los límites del derecho de acceso a la información pública, particularmente, sobre una aplicación informática de toma de decisiones automatizadas, que esta Sala llevará a cabo a continuación, se encuentra condicionada al hecho de que ese funcionamiento automatizado sirve de soporte al reconocimiento o denegación de derechos sociales, arrojando un resultado positivo o negativo, sin exteriorizar las razones de dicho resultado.

Por lo demás, en estos casos la transparencia de las aplicaciones informáticas o del proceso tecnológico seguido por el sistema informático adquiere singular relevancia para garantizar el adecuado control de la gestión pública, al brindar a la ciudadanía la información necesaria acerca del proceso seguido en la toma de decisiones para su comprensión, así como para comprobar su adecuación a las normas cuya aplicación debe regir su funcionamiento.

Realizadas las anteriores consideraciones generales, estamos en condiciones de abordar la ponderación de los intereses relativos a los límites al acceso a la información pública alegados.

Con carácter general, en la ponderación de intereses en juego se tiene en cuenta las siguientes circunstancias sobre el derecho de acceso ejercitado por la fundación Civio:

- Se concreta por la parte demandante en un interés público en acceder al código fuente de la aplicación BOSCO para conocer cómo se toma la decisión de conceder o rechazar el bono social y comprobar si existen errores en dicha decisión. Interés que se pone en relación con el derecho a la seguridad jurídica de las personas administradas, ya que la aplicación BOSCO es empleada para el reconocimiento de derechos a los consumidores (se han detectado errores en la denegación de la asignación de esa condición a viudas, y en la exigencia en caso de familias numerosas de consentimiento informado no previsto normativamente)
- Resulta relevante considerar que el bono social, que se materializa en un descuento en la factura de consumo de energía eléctrica, se inserta entre las medidas previstas por la Ley 24/2013, de 26 de diciembre, del Sector Eléctrico (artículos 45 y 45 bis) para proteger a los consumidores vulnerables y luchar contra la pobreza energética.
- Relevancia pública de la información a la que se pretende el acceso, pues el código fuente de dicha aplicación ha de responder a las disposiciones normativas que regulan los requisitos que deben cumplir los consumidores para el reconocimiento del bono social por ostentar la condición jurídica de consumidor vulnerable,
- La Fundación Ciudadana Civio, que es una organización independiente y sin ánimo de lucro, cuya actividad responde a la vigilancia del funcionamiento de las instituciones públicas y de la gestión de los recursos públicos que llevan a cabo, así como la promoción de la información de los ciudadanos acerca de su funcionamiento, para lo cual persigue fomentar su transparencia.
- El legítimo interés de la fundación recurrente en acceder al código fuente de la aplicación informática reside en la verificación del correcto funcionamiento de la aplicación telemática BOSCO, contrastando que la aplicación telemática es fiel a las previsiones normativas que establecen los requisitos necesarios para ser considerado consumidor vulnerable.
- La información objeto de acceso proporciona transparencia sobre los asuntos públicos y es relevante para la sociedad en su conjunto o, al menos, para una parte especialmente débil de la misma -los consumidores que se encuentran en una situación social y económica más frágil frente a la pobreza energética-, lo que evidencia su interés público, con independencia de que la Fundación Ciudadana Civio sea una entidad privada.

- El programa BOSCO toma una decisión que condiciona el acceso al bono social eléctrico, que no viene acompañada de la expresión de los motivos concretos que sustentan dicha conclusión.

1.- El límite al acceso a la información pública consistente en la <u>propiedad intelectual</u> del artículo 14.1.j) de la LTAIBG.

La aplicación BOSCO es un programa de ordenador, incluido específicamente dentro de las creaciones susceptibles de ser objeto de propiedad intelectual (artículos 10.1.i) y 95 y siguientes del TRLPI).

La sentencia descarta la aplicación a BOSCO del artículo 13 del Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual (TRLPI) porque BOSCO sea un acto administrativo o una disposición general: *«No son objeto de propiedad intelectual las disposiciones legales o reglamentarias y sus correspondientes proyectos, las resoluciones de los órganos jurisdiccionales y los actos, acuerdos, deliberaciones y dictámenes de los organismos públicos, así como las traducciones oficiales de todos los textos anteriores».*

Existe un debate en la doctrina sobre la naturaleza jurídica de los algoritmos, que desarrollaremos más adelante, y en algún comentario a la sentencia se ha sostenido (Ponce Solé) que BOSCO no es una mera aplicación instrumental o auxiliar, sino un sistema de IA que adopta decisiones totalmente automatizadas, el cual supone la traducción a código informático de una norma previa, para su aplicación digital en el futuro en infinidad de ocasiones, hubiera sido posible optar por la consideración de este como disposición general. Entiende que no todos los algoritmos son siempre reglamentos, pero sí lo son aquellos que desarrollan una ley al traducirla a lenguaje formal para su lectura por máquinas.

Se estima oponible el límite del artículo 14.1.j) de la LTAIBG -propiedad intelectual-, pero el mero riesgo de eventuales perjuicios para la Administración titular del derecho de propiedad intelectual concernido con motivo de su uso o explotación no autorizada, como consecuencia del acceso a la información pública -acceso al código fuente-, por sí solo, no constituye una causa de exclusión del derecho de acceso, al carecer de la relevancia necesaria para operar con tal efecto limitativo, considerando en la ponderación de intereses que:

- La protección jurídica que proporciona la propiedad intelectual (en relación a las facultades patrimoniales que integran el derecho de autor) y la finalidad que persigue se ve atenuada cuando, como sucede aquí, el programa de ordenador ha sido creado por la propia Administración Pública, que es la titular de la propiedad intelectual, por mandato de la normativa del sector eléctrico para el ejercicio de competencias públicas y dirigida a servir a intereses igualmente públicos, no encontrándose, en consecuencia, integrada -o no, al menos, principalmente- en la lógica competitiva del mercado donde se proyectan con especial significación los derechos de explotación de la propiedad intelectual.

- Los eventuales perjuicios que pudieran dimanar de dicho acceso como consecuencia de su explotación no autorizada por terceros; perjuicios cuyo riesgo fácilmente puede ser minimizado sometiendo el acceso a determinadas cautelas, como, por ejemplo, la prohibición de la difusión o la utilización del código fuente para otras finalidades sin la autorización expresa de la Administración, la advertencia expresa de la responsabilidad en que puede incurrir el solicitante de acceso por el incumplimiento de esa prohibición, la firma de un compromiso de uso limitado de la información recibida o la imposición de un deber de reserva o confidencialidad respecto de la información consultada.

En conclusión, se en la ponderación de intereses en juego que hace esta Sala, se otorga prevalencia al interés en el acceso al código fuente de la aplicación telemática BOSCO sobre el derecho a la propiedad intelectual de la Administración del Estado.

2.- El límite al acceso a la información pública consistente en la <u>seguridad pública</u> del artículo 14.1.d) de la LTAIBG.

Partiendo de que, en general, la revelación del código fuente aumenta de una manera objetiva la severidad de las vulnerabilidades de cualquier aplicación informática, tal riesgo *per se* no puede oponerse, sin más, al derecho de acceso a la información pública que entraña el algoritmo. De modo que el límite de seguridad pública solo puede impedir el acceso a la información pública cuando ello se encuentre justificado y resulte proporcionado a su objeto y finalidad, previa ponderación de los intereses en conflicto, ex artículo 14.2 de la LTAIBG.

En esa ponderación de intereses resulta relevante el grado de vulnerabilidad inherente a aquel acceso y el peligro que pudiera suponer para el acceso no consentido a los datos personales de solicitantes del bono social, por un lado, y la relevancia pública de la información y el legítimo interés de la fundación al acceso, por otro lado.

En esta ponderación de derechos e intereses no pueden desdeñarse los riesgos de seguridad que pudiera generar el acceso de terceros al código fuente del algoritmo del sistema informático por las vulnerabilidades que entrañe. Pero tampoco puede soslayarse que estos riesgos, por lo general, pueden ser previstos, lo que posibilita el diseño de la aplicación o programa informático fortaleciendo la seguridad del sistema, con su consiguiente minimización.

Aun cuando existen evidentes intereses y derechos relacionados con la confidencialidad, la protección de datos personales y la seguridad informática que deben quedar preferentemente tutelados cuando las circunstancias específicas de cada caso así lo aconsejen, es apreciable que, tanto en la normativa de la Unión Europea, como en la normativa doméstica existen mandatos y principios favorables a la transparencia de los algoritmos públicos

que conducen a descartar la ocultación del código fuente como principio general y categórico de seguridad de los sistemas informáticos.

Todas estas circunstancias justifican la prevalencia en este caso del interés público esgrimido por la fundación solicitante de acceder el código fuente del programa informático frente al límite del artículo 14.1.d) de la LTAIBG opuesto por la Administración recurrida, habida cuenta de que los riesgos de seguridad, además de no quedar suficientemente caracterizados, se verían en todo caso circunscritos a una operativa informática especialmente acotada (la de la aplicación del bono social eléctrico), mientras que los intereses relativos al control de la actuación de la Administración, conectados con la significativa finalidad para la que se emplea el programa BOSCO, y, por ende, el correcto funcionamiento del programa informático, sirven, en este caso, a la efectividad de relevantes bienes jurídicos como los principios de legalidad e igualdad y otros derechos constitucionales, la participación en la toma de decisiones, el fortalecimiento de la democracia y la generación de confianza en las instituciones públicas.

En el caso examinado no se había justificado la concurrencia de los límites previstos en letras g), i) y k) del artículo 14.1 de la LTAIBG -las funciones administrativas de vigilancia, inspección y control, la política económica y monetaria y la garantía de la confidencialidad o el secreto requerido en procesos de toma de decisión-.

Como consecuencia de todo lo hasta aquí expuesto, considera el Tribunal Supremo que la transparencia exigible en el funcionamiento de la aplicación telemática BOSCO, con evidente y relevante impacto en los derechos sociales de los ciudadanos, no queda garantizada con la mera explicación funcional sobre la misma, ofrecida por la Administración titular de la aplicación, sino que exige el acceso a su algoritmo, pues de otro modo no resultaría posible comprobar con exactitud y detalle dicho funcionamiento y, por ende, la sujeción de las ordenes o instrucciones en lenguaje informático que contiene a las previsiones legales y reglamentarias sobre los requisitos necesarios para la obtención del bono social.

En el caso enjuiciado la Fundación Ciudadana Civio desempeña funciones de vigilancia social asociadas a la guarda y custodia del Estado de Derecho y, por ende, de la democracia, en la medida que pretende velar por el correcto funcionamiento de las instituciones públicas y promover la información de los ciudadanos acerca de su mismo y la gestión de los recursos públicos.

Por tanto, se cumplían los presupuestos que el Tribunal Europeo de Derechos Humanos ha exigido para reconocer el derecho de acceso a información pública, que expone en su Sentencia 18030/11, de 8 de noviembre de 2016 (asunto Maygar Helsinki Bizottság c. Hungría): (i) propósito del acceso -la finalidad buscada debe generar debate público o contribuir a la labor de vigilancia de los poderes públicos-; (ii) naturaleza de la información -la información debe ser de interés público, es decir, que proporcione transparencia sobre los asuntos públicos o sea relevante para la sociedad en su conjunto-; (iii) rol del solicitante -debe tratarse de un solicitante que desempeñe funciones de vigilancia social asociadas a la guardia y custodia de la democracia (el TEDH

entiende que cumplen estas condiciones los medios de comunicación y periodistas, pero también organizaciones no gubernamentales, investigadores académicos, escritores sobre materias de interés público, blogueros, influencers, etc.)-, y (iv) existencia y disposición de la información -la información debe estar disponible, sin que sea necesario acometer una labor de recopilación-.

Repárese en que la LTAIBG reconoce el derecho de acceso a la información pública con mayor amplitud que el TEDH, pues lo hace también en base a solicitudes fundadas en intereses privados legítimos, por lo que no delimita negativamente el ámbito subjetivo del derecho de acceso por razón del interés privado que lo motive, ni tampoco exige acreditar un determinado interés, tal y como se deduce de su artículo 12 y reconoce nuestra jurisprudencia [STS de 12 de noviembre de 2020 (rec. 5239/2019) y de 2 de junio de 2022 (rec. 4116/2020)], con independencia de que pueda tomarse en consideración en la ponderación de los bienes jurídicos confrontados.

D/ Jurisprudencia fijada

De conformidad con las consideraciones expuestas en los apartados anteriores, esta Sala, dando respuesta a la cuestión de interés casacional planteada en este recurso de casación, que presenta interés casacional para la formación de jurisprudencia, en interpretación y aplicación de los artículos 14 y 16 de la Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno, a la luz del artículo 42 de la Carta de Derechos Fundamentales de la Unión Europea y del artículo 105.b) de la Constitución Española, declara lo siguiente:

1.- El derecho de acceso a la información pública trasciende a su condición de principio objetivo rector de la actuación de las Administraciones públicas, para constituir un derecho constitucional ejercitable, como derecho subjetivo, frente a las Administraciones públicas, derivado de exigencias de democracia y transparencia, e inseparablemente unido al Estado democrático y de Derecho.

2.- El derecho de acceso a la información pública adquiere especial relevancia ante los riesgos que entraña el uso de las nuevas tecnologías en el ejercicio de las potestades públicas o la prestación de servicios públicos, como ocurre con el empleo de sistemas informáticos de toma de decisiones automatizadas en la actividad de las Administraciones públicas, especialmente, cuando tienen por objeto el reconocimiento de derechos sociales. En estos casos debe conllevar exigencias de transparencia de los procesos informáticos seguidos en dichas actuaciones, con el objeto de proporcionar a los ciudadanos la información necesaria para su comprensión y el conocimiento de su funcionamiento, lo que puede requerir, en ocasiones, el acceso a su código fuente, a fin de posibilitar la comprobación de la conformidad del sistema algorítmico con las previsiones normativas que debe aplicar.

3.- La Fundación Ciudadana Civio tiene derecho a acceder al código fuente de la aplicación informática BOSCO, desarrollada para que las empresas comercializadoras de referencia de energía eléctrica puedan comprobar si los solicitantes del bono social cumplen con los requisitos previstos, legal y

reglamentariamente, para tener la consideración de consumidor vulnerable y, por ende, resultan ser beneficiarios del bono social, con la finalidad de que pueda conocer las operaciones diseñadas para la concesión del bono social y comprobar que se ajustan al marco normativo aplicable.

E/ Consideraciones finales.

La sentencia del Caso BOSCO se dicta cuando la aplicación de sistemas de inteligencia artificial en la actividad administrativa resulta aún incipiente, pero avanza aceleradamente. Su doctrina sobre transparencia algorítmica resulta directamente aplicable —y adquirirá especial relevancia— cuando las Administraciones empleen sistemas de IA, especialmente algoritmos de aprendizaje automático o redes neuronales.

Estos sistemas presentan el desafío adicional de la opacidad intrínseca: incluso accediendo al código fuente, resulta difícil explicar por qué un sistema de *deep learning* adopta determinada decisión, dada la complejidad de las capas de abstracción y la imposibilidad de seguir linealmente el proceso decisorio. Pero precisamente por ello, la exigencia de transparencia algorítmica resulta más imperiosa en esos casos: las Administraciones no podrán emplear sistemas de IA tipo «caja negra» para adoptar decisiones sobre derechos ciudadanos sin garantizar mecanismos de explicabilidad efectiva.

La doctrina del Caso BOSCO anticipa así los debates que generará la aplicación del RIA, pues las Administraciones que pretendan emplear IA en sus procesos decisorios deberán diseñar desde el origen sistemas auditables, trazables y explicables, bajo el presupuesto de su futura escrutabilidad pública.

Desde la perspectiva de esta sentencia, si la Administración desea hacer uso de aplicaciones de decisiones administrativas automatizadas, debe cumplir exigencias de transparencia y ser capaz de ofrecer servicios con algoritmos no opacos, que expliquen las decisiones que toman, y que puedan ser verificables por entidades expertas independientes para comprobar que trasladan correctamente las decisiones del legislador.

De lo contrario se exponen al acceso público a toda la documentación que tengan para que se verifique la correcta traslación de la norma al algoritmo, lo que incluye, si es preciso, el código fuente, con independencia de que ello afecte a su derecho de propiedad intelectual o si pudiera causar un eventual efecto, no contrastado, en la ciberseguridad.

No obstante, como hemos razonado, debe reconocerse que el acceso al código fuente no es la única, ni la mejor, respuesta a la transparencia, pues al margen de los límites legales que establece la propia LTAIBG, cuya aplicación puede obstaculizar dicho acceso, resulta incomprensible para la mayoría de los ciudadanos y su publicidad puede impedir a las Administraciones el acceso a mercados de proveedores tecnológicos cualificados. Por ello, sin perjuicio de que

deba impulsase la utilización por las Administraciones públicas de aplicaciones de código abierto, cuando ello no sea oportuno y la satisfacción de los intereses generales aconseje el empleo de aplicaciones en las que no resulte viable el acceso al código fuente, deberán implementarse otras medidas para cumplir las exigencias de transparencia.

La sentencia apunta también a la necesidad de asegurar la intervención de entidades independientes de control de las tecnologías de la información y comunicación (TIC) para garantizar que la IA empleada es fiable y segura.

La Agencia Española de Supervisión de Inteligencia Artificial (AESIA) podría cumplir esa función de evaluación y control de los algoritmos públicos. Su creación fue autorizada primero por la Ley 22/2021, de 28 de diciembre, de Presupuestos Generales del Estado para el año 2022 (disposición adicional centésima trigésima) y después con mayor densidad normativa por la Ley 28/2022, de 21 de diciembre, de fomento del ecosistema de empresas emergentes (Disposición adicional séptima) para dar cumplimiento a las previsiones del Reglamento (UE) 2024/1689, como autoridad nacional encargada de supervisar su aplicación, coordinar las actuaciones de los Estados miembros, actuar como punto de contacto único con la Comisión Europea y representar al Estado español ante el Comité Europeo de Inteligencia Artificial.

En efecto, la Agencia Española de Supervisión de Inteligencia Artificial (AESIA), tiene, entre otros fines, la *«supervisión de la puesta en marcha, uso o comercialización de sistemas que incluyan inteligencia artificial y, especialmente, aquellos que puedan suponer riesgos significativos para la salud, seguridad y los derechos fundamentales»,* sin distinguir entre usos de Derecho Público y usos de Derecho Privado [Disposición adicional séptima de la Ley 28/2022, Dos, e)].

El Real Decreto 729/2023, de 22 de agosto, por el que se aprueba el Estatuto de la Agencia Española de Supervisión de la Inteligencia Artificial, detalla sus funciones y, entre otras, la Subdirección de Certificación, Evaluación de Tendencias, Coordinación y Formación en inteligencia artificial tiene atribuidas las siguientes competencias: (1) Supervisar de oficio los sistemas de inteligencia artificial utilizados por las administraciones públicas, así como emitir informes vinculantes y actas que decidan sobre la continuidad de dichos sistemas y/o su puesta en marcha (art. 26.a.6º); y (2) Apoyar técnicamente y/o asesorar a jueces y tribunales en casos de conflictos relacionados con la inteligencia artificial, a requerimiento de los propios tribunales (art. 26.d.4º). En este sentido, la AESIA podría intervenir en procesos judiciales para proporcionar un informe a instancias del órgano jurisdiccional, lo que resulta especialmente relevante dado el alto coste de las periciales en los ámbitos tecnológicamente complejos.

La AESIA tiene sede en Coruña y recientemente ha publicado unas primeras guías para orientar la aplicación del RIA, que recoge diversas obligaciones de transparencia a lo largo de su articulado. En particular, interesa a los presentes efectos la Guía 8: Transparencia y provisión de

información a los usuarios, cuya primera versión es de 10 de diciembre de 2025 y que se encuentra disponible en el sitio internet de la Agencia.