derecom

ISSN: 1988-2629. Nº. 31. Nueva Época. Septiembre 2021-Marzo 2022

Cómo citar este texto:
Gutiérrez David, M.E. (2021). Administraciones inteligentes y acceso al código fuente y los algoritmos públicos. Conjurando riesgos de cajas negras decisionales, *Derecom*, 31, 19-105, http://www.derecom.com/derecom/

ADMINISTRACIONES INTELIGENTES Y ACCESO AL CÓDIGO FUENTE Y LOS ALGORITMOS PÚBLICOS. CONJURANDO RIESGOS DE CAJAS NEGRAS DECISIONALES[1]

SMART GOVERNMENTS AND ACCESS TO SOURCE CODE AND ALGORITHMS HELD BY PUBLIC AUTHORITIES. DISMISSING RISKS OF A BLACK BOX DECISION-LIKE[2]

© María Estrella Gutiérrez David
Universidad Complutense de Madrid (España)
esgutier@ucm.es

Resumen

Existe un consenso amplio en que el uso de sistemas de IA por parte de las Administraciones públicas debe ser transparente y garantizar que los ciudadanos comprenden cómo y por qué se han adoptado las decisiones algorítmicas que les afectan individual o colectivamente. La legislación de transparencia puede y debe ser un instrumento para este propósito. Existe una doctrina bien asentada de las Autoridades de transparencia que considera que código fuente y los algoritmos (deterministas o predictivos) utilizados por la Administración constituye «información pública». Partiendo de esta premisa, la casuística comparada e interna del derecho de acceso a la información posibilita identificar algunos de los riesgos inherentes al uso de estos sistemas de IA por parte de las Administraciones: reglamentación oculta y errores; existencia de black boxes decisionales; sesgos embebidos y vulneraciones de derechos y libertades. Sin perjuicio de los eventuales límites legales que puedan concurrir (e.g. seguridad pública, propiedad intelectual e industrial), analizados en este trabajo, existe cierto consenso en que el derecho de acceso no siempre garantiza una total transparencia y comprensibilidad del proceso de toma de decisiones mediante sistemas de IA, especialmente, cuando estos utilizan modelos de black box. Es más, la noción misma de un black box como incapacidad de entender los resultados producidos por un sistema algorítmico comienza a ser expansiva, en el sentido de que el análisis de la casuística ahí donde el acceso al código fuente de algoritmos deterministas es objeto de litigio evidencia que ni los afectados por la decisión ni el juez son capaces de entender cómo el sistema llegó a tal decisión. Se constata, por último, que no existe una coincidencia exacta entre el significado técnico de la transparencia algorítmica manejado por la XAI y el significado jurídico de la transparencia administrativa propio del iuspublicismo. De lege ferenda, y a fin de garantizar una adecuada interpretabilidad, explicabilidad y justificación de las decisiones públicas adoptadas o apoyadas en sistemas de IA y el escrutinio público de las mismas, la legislación de transparencia debería determinar la información relevante que podría ser objeto de publicidad activa o del derecho de acceso e incorporar obligaciones de producción, conservación y registro de la documentación técnica relativa al ciclo de vida de tales sistemas.

Recibido: 08/02/2021 Revisado: 24/02/2021
Aceptado: 04/02/2021 Publicado:15/09/2021

Summary

There is broad consensus that the use of AI systems by Governments must be transparent and ensure that citizens are able to understand how and why algorithmic decisions affecting them, individually or collectively, have been made. Transparency legislation can be a useful instrument for this purpose. There is a well-established doctrine issued by freedom information Authorities that the source code and algorithms (either deterministic or predictive) used by the Governments are «public information». Based on that premise, the comparative and domestic casuistic of the right of access to public information makes possible to identify some of the risks inherent in the use of AI systems by Governments: covert regulation and bugs; existence of black box decision-like; embedded biases and impairment of rights and freedoms. Notwithstanding the potential legal limits (e.g. public security, intellectual property) analysed in this paper, there is a wide consensus that the right of access does not always ensure full transparency and understanding of the algorithmic decision-making process, especially where black box models are implemented. Moreover, the idea itself of black box as an inability to understand how an algorithmic system produced an output is becoming expansive. In effect, cases where requests of access concerning the source code of deterministic algorithms is at dispute are showing to what extent neither the affected parties by an algorithmic decision nor the judge are able to understand how the system reached such decision. In addition, it has been found that there is no an exact match between the technical meaning of «algorithmic transparency» handled by the «XAI» and the legal meaning of «administrative transparency» typical of Public Law. On a lex ferenda basis, in order to achieve appropriate interpretability, explainability and justification for governmental decisions made or supported by AI systems and public scrutiny thereof, freedom of information legislation should determine relevant information to be publicly disclosed or accessed and ensure that technical documents related to the whole life-cycle of such AI systems are dully produced, kept and registered.

Palabras clave: Derecho de acceso. Código fuente. Algoritmos de aprendizaje automatizado. Black box decisional. Transparencia algorítmica. Interpretabilidad. Explicabilidad.

Keywords: Freedom of information. Source code. Machine learning algorithms. Black-box decisions-like. Algorithmic transparency. Interpretability. Explainability.

1.Introducción

Las llamadas *tecnologías habilitadoras clave*, como la Inteligencia Artificial (IA), el Big Data, Internet de las Cosas (IoT), las redes 5G, *Blockchain*, o la Computación Cuántica permiten mayores capacidades de procesamiento de la información garantizando su recuperación, conservación lógica y trazabilidad. En particular, las Administraciones y el sector público no son ajenos a la incorporación de estas tecnologías de impactos sociales disruptivos, particularmente, el Big Data y la IA, en sus modalidades *Machine Learning* (ML) y *Deep Learning* (DL).

Pero la realidad es que buena parte de estos tratamientos mediante sistemas de IA, a menudo, están ocultos para los afectados y para la sociedad en general. A la opacidad inherente de estos sistemas se han referido la Autoridad de Transparencia y Protección de Datos del Reino

Unido, el *Information Commissioner's Office* (ICO), como *tratamientos invisibles* (*invisible processing*).[3] De hecho, al no existir en las legislaciones de transparencia obligaciones de publicidad activa respecto de los sistemas de IA implementados por el sector público no es posible realizar un mapa fiable y completo de los usos de estos sistemas y su alcance real.

En su descripción de la *Black Box Society*, Pascal señala que *la transparencia no es sólo un fin en sí mismo, sino un paso intermedio en el camino hacia la inteligibilidad*.[4] La afirmación es pertinente a la hora de entender las limitaciones de la legislación de transparencia y, en consecuencia, del derecho de acceso, para garantizar el escrutinio público de las decisiones administrativas basadas o adoptadas mediante la IA.

1.1. Estado de la cuestión. El interés iusfundamental del tema desde la perspectiva del derecho de acceso a la información pública

En el marco de Naciones Unidas, la Agenda 2030 incluye, entre los *objetivos de desarrollo sostenible* la necesaria promoción de sociedades pacíficas e inclusivas (ODS número 16). Ello incluye la implantación efectiva del Estado de Derecho en los planos nacional e internacional, garantizando la igualdad de acceso a la justicia para todos (16.3), el desarrollo de instituciones eficaces, responsables y transparentes a todos los niveles (16.6), asegurar procesos de toma de decisiones receptivos, inclusivos, participativos y representativos en todos los niveles (16.7) y garantizar el acceso público a la información y proteger las libertades fundamentales, de conformidad con la legislación nacional y los acuerdos internacionales (16.10):[5]

Los objetivos específicos que acaban de identificarse de la Agenda 2030 deben enmarcarse, sin duda, en la cada vez mayor preocupación política y social por la creciente opacidad en el uso de sistemas de IA en la actividad ordinaria y procesos de toma de decisiones de organizaciones públicas y privadas con impactos relevantes en los derechos humanos. El Consejo de Europa ha subrayado así la importancia de la transparencia como instrumento de escrutinio público de estos sistemas, especialmente cuando son utilizados en el contexto de la prestación de servicios públicos.

> *El uso de un sistema de IA en cualquier proceso de toma de decisiones* –dice el Consejo de Europa– *con un impacto significativo en los derechos humanos de las personas necesita ser identificable. El uso de un sistema de IA no solo deber ser dado a conocer públicamente en términos claros y accesibles, sino que las personas individuales deben ser capaces de comprender cómo se adoptan las decisiones y cómo se han verificado esas decisiones*.

En este sentido, la imposición de obligaciones de transparencia posibilitaría la supervisión de los sistemas de inteligencia artificial. A juicio del Consejo de Europa, tales obligaciones de transparencia –y esto es lo relevante– pueden consistir en *la divulgación pública de información sobre el sistema en cuestión, sus procesos, los efectos directos e indirectos sobre*

los derechos humanos, y las medidas tomadas para identificar y mitigar los impactos adversos del sistema en los derechos humanos.[6]

En consecuencia, el despliegue y uso de sistemas algorítmicos de IA por parte de las organizaciones públicas y privadas debe ser abordado jurídicamente con extremo cuidado en la medida en que los derechos fundamentales puedan estar particularmente afectados. Siendo esto así, el Comisionado Alemán de Ética de los Datos considera que las decisiones adoptadas mediante sistemas algorítmicos deben ser transparentes y justificables, y este deber adquiere una relevancia en el sector público si cabe aún mayor que en el privado, en tanto en cuanto las Administraciones públicas están sujetas al principio de legalidad y vinculación positiva al Estado de Derecho, lo que incluye el deber inexcusable de respetar los derechos fundamentales y la necesidad democrática de rendición de cuentas. En consecuencia, a juicio del Comisionado, *no solo deben aplicarse requisitos generales de transparencia a las entidades públicas, sino que además éstas deben esforzarse particularmente en garantizar su apertura*. En este sentido, en la mayoría de los casos, los sistemas algorítmicos utilizados por el sector público *entran en el ámbito de aplicación del vigente derecho de acceso a la información pública y/o la legislación de transparencia*.[7]

En este contexto, no puede pasar desapercibida la *Enmienda Belot* al Proyecto de Ley de República Digital por la cual se incluyó el *código fuente* en la enumeración de documentos administrativos sujetos al derecho de acceso, en la cual el autor de dicha enmienda evidenciaba cómo las solicitudes de acceso al código fuente utilizado por una Administración y *que hace intervenir a los algoritmos* en la toma de decisiones individuales se han convertido en una "cuestión recurrente"[8]. Las resoluciones previas de la Commission d'Accès aux Documents Administratifs (CADA), la Autoridad francesa de transparencia, con relación a solicitudes cuyo objeto era el acceso al código fuente o al algoritmo decisional subyacente, así lo acreditaban.

El análisis de la doctrina administrativa y de la jurisprudencia, especialmente la comparada, evidencia que cada vez son más frecuentes las solicitudes que tienen por objeto el acceso al código fuente de las aplicaciones informáticas utilizadas por la Administración y a los modelos algorítmicos subyacentes en las mismas.

En un interesante estudio realizado por Katherine Fink, la autora analiza las solicitudes presentadas al amparo de la FOIA entre 2012 y 2016 y que tuvieran por objeto el acceso al *código fuente* o a los algoritmos implementados por las agencias federales. Dicho estudio identificó hasta 73 solicitudes, de las cuales 21 fueron estimatorias, otras 30 fueron denegadas y otras 22 estaban aún pendientes de contestación o su resolución era desconocida.[9]

A pesar de las reticencias de las Administraciones a facilitar el acceso a estos activos digitales de información en su poder, el análisis de la doctrina comparada pone de manifiesto que no sólo empieza a reconocerse su carácter de información pública, con independencia del soporte o lenguaje en que estén expresados, sino también el interés público en el acceso a dicha información al posibilitar el control de las decisiones públicas y cómo han sido adoptadas.

1.2. Objeto de estudio, hipótesis y metodología

Habida cuenta el contexto descrito, el objeto de estudio del presente artículo pretende abordar en qué medida el derecho de acceso al código fuente de las aplicaciones utilizadas por la Administración y, en su caso, a los algoritmos subyacentes, puede constituir un instrumento de control de la actividad, formalizada o no, de una Administración «inteligente», caracterizada por

una creciente automatización, datificación y algoritmización, así como de los eventuales impactos de tal actividad en los derechos y libertades públicas.

En particular, el artículo centrará su atención en el derecho de acceso a los algoritmos de IA implementados por las Administraciones y sector público, y en qué medida el ejercicio de este derecho, o en su caso, la existencia de obligaciones de publicidad activa, pueden garantizar un adecuado escrutinio público de los sistemas de IA utilizados por la Administraciones en su proceso de toma de decisiones.

Nuestro objeto de estudio así definido parte de las siguientes hipótesis de trabajo.

En primer lugar, los activos de información digital aquí identificados (código fuente y modelo o modelos algorítmicos subyacentes) y en poder de los sujetos obligados por la legislación de transparencia, deben considerarse información pública.

En segundo lugar, el derecho de acceso puede constituir un instrumento más para garantizar la transparencia algorítmica en particular, y la transparencia de la Administración en general.

En tercer lugar, existe un interés público en garantizar el derecho de acceso a estos activos de información pública digital, en la medida en que permiten la trazabilidad y el control de las decisiones adoptadas por la Administración y responsables públicos en un contexto de automatización, datificación y algoritmización creciente con evidente impacto en los derechos y libertades del ciudadano.

En cuarto lugar, la existencia de dicho interés público en el acceso no empece la debida ponderación con otros bienes jurídicos dignos también de protección amparados por la legislación de transparencia, así como la legislación general o sectorial aplicable, en particular, la seguridad pública o los derechos de propiedad intelectual e industrial.

En quinto lugar, la casuística administrativa y judicial analizada sobre el derecho de acceso evidencia la existencia de una serie de riesgos en la adopción de sistemas automatizados o semi-automatizados de toma de decisiones en el ámbito público, al margen de la naturaleza del algoritmo implementado, determinista o de IA. De hecho, en buena parte de los conflictos en los que se reclama el acceso al código fuente del algoritmo utilizado por una Administración no se discute la naturaleza del algoritmo en cuestión (si es determinista, de aprendizaje automatizado o si, dentro de esta última categoría, es un modelo de *black box*), sino en qué medida la decisión adoptada con base en, o mediante, el sistema automatizado ha resultado comprensible o no para los afectados.

En sexto lugar, con relación a los sistemas algorítmicos de IA implementados por la Administración el derecho de acceso al código fuente o a la documentación técnica del algoritmo no siempre permitirá entender al ciudadano (no experto) el proceso de toma de decisiones, particularmente, cuando se trate de modelos de caja negra, aunque no exclusivamente. La legislación de transparencia actual tiene limitaciones que deben ser corregidas para garantizar el adecuado escrutinio público de la algoritmia decisional en el ámbito de las Administraciones y el sector público.

Para abordar nuestro objeto de estudio y confirmar nuestras hipótesis de trabajo, metodológicamente se ha optado por el análisis y sistematización de la doctrina emanada de las Autoridades de Transparencia y de los Tribunales de Justicia –especialmente la del orden contencioso-administrativo, tanto en el ámbito comparado como el nacional.

Así, por ejemplo, desde que en 2016 la CADA reconociese el carácter de documento administrativo al código fuente y a los algoritmos utilizados por las Administraciones, la Autoridad francesa ha dictado 17 resoluciones en total a la fecha de redacción de este artículo. De las 7 resoluciones relativas conjuntamente al derecho de acceso al código fuente y a los tratamientos algorítmpicos implementados por aplicaciones informáticas utilizadas por las Administraciones, 5 son estimatorias y 2 desestimatorias. Asimismo, constan 6 resoluciones relativas al acceso al código exclusivamente, de las cuales 3 son estimatorias y otras 3 desestimatorias, así como 4 resoluciones que estiman el derecho del interesado a una explicación sobre los tratamientos algorítmicos que sean fundamento de decisiones individuales.[10] Pues bien, tal doctrina es objeto de sistematización y análisis en este trabajo.

A partir del examen casuístico puede avanzarse ya que en el ámbito comparado comienzan a establecer líneas interpretativas que, sin duda, podrían calificarse *pro informatione*, donde debe destacarse, por ejemplo, la línea abierta por la CADA en Francia o el Tribunal Supremo de los Países Bajos al reconocer el derecho de acceso al código fuente y/o al modelo algorítmico subyacente.

Por contra, debe subrayarse que los asuntos abordados en nuestro Derecho interno en aplicación de la Ley 19/2013, de 9 de diciembre, de Transparencia, Acceso a la Información Pública y Buen Gobierno *(LTAIBG*) son escasos, y los criterios mantenidos por las Autoridades españolas de transparencia, fundamentalmente, el Consejo de Transparencia y Buen Gobierno estatal (CTBG) y la Comisión Catalana de Garantías de Acceso a la Información Pública (GAIP), son dispares, lo cual dificulta la sistematización de una doctrina sólida y coherente respecto al tratamiento de este tipo de solicitudes de acceso.

En nuestra aproximación metodológica al objeto de estudio también resulta decisiva la aportación de la doctrina científica comparada en la cual debe destacarse el esfuerzo intelectual bidireccional realizado desde el campo de las Ciencias Exactas para divulgar los entresijos de estas tecnologías disruptivas *nivel de hombre de la calle*, así como desde el campo del Derecho por aprehender el significado técnico de los conceptos vinculados a estas tecnologías aquí referidas, a las matemáticas y a la programación, para traducirlos después en categorías jurídicas. La importancia de este esfuerzo intelectual bidireccional no puede pasar desapercibida.

Difícilmente puede garantizarse la corrección, validez y justiciabilidad de las decisiones producidas o asistidas por esta clase de modelos si los operadores jurídicos, como la Administración o el Juez desconocen cuál es la taxonomía de datos que conforma el conjunto de entrenamiento o cómo se han preparado esos datos (eg. tratamiento de datos desequilibrados, faltantes, de cardinalidad alta), qué variables han sido determinantes en el resultado y por qué; cuáles han sido los parámetros resultantes (e.g. los *pesos* en una red neuronal), los hiperparámetros seleccionados y por qué (e.g. la tasa de aprendizaje o el número de árboles que componen un bosque aleatorio); qué es la *generalización* y el *overfitting*; en qué consiste la *maldición de la dimensionalidad*; cómo se analizan los errores de un modelo, cómo

se valida y testea; o qué técnicas suplementarias pueden aplicarse para posibilitar la interpretabilidad de un modelo de *black box*. La complacencia con los resultados del sistema implementado, por muy *inteligente* que sea no puede ser la solución, especialmente, si los resultados son lesivos para los derechos fundamentales de los ciudadanos.

En efecto, el legislador debe conocer y comprender los fundamentos básicos de estas tecnologías pues debe regular adecuadamente esta realidad introduciendo las garantías adecuadas frente a los riesgos de estas tecnologías, reconociendo, si es preciso, nuevos derechos o extendiendo el ámbito de ejercicio de otros ya existentes. Por su parte, la Administración también debe conocer y comprender ya que cada vez está más inmersa en la adquisición de estas tecnologías avanzadas dejando, en muchos casos, libertad al contratista para que diseñe el modelo sin evaluaciones de impacto previas. Y, en fin, en su control de legalidad de las decisiones adoptadas con estas tecnologías el Juez igualmente debe conocer y comprender la racionalidad subyacente en estos modelos. Que el modelo sea un *black box* o complejo desde el punto de vista computacional no puede ser la excusa para seguir manteniendo la opacidad del uso de estos sistemas en el ámbito público.

2. Automatización, inteligencia y disruptividad en nuestras Administraciones: trazabilidad de la IA en el ámbito público

Más allá de la implantación progresiva de la e-Administración, las entidades públicas sujetas a la legislación de transparencia y acceso a la información pública, y muy particularmente, las Administraciones, no son ajenas al impacto de las llamadas *tecnologías disruptivas digitales*, en particular el Big Data y la IA.

Es importante tener claro que, cuando hablamos de la aplicación de analítica Big Data y de la IA a la actividad administrativa, formalizada o material, nos referimos al tratamiento masivo de datos, personales o no, por medios automatizados o semi-automatizados.

Los procesos de analítica Big Data están asociados al tratamiento de datos *a gran escala*, y se vienen caracterizando habitualmente por cuatro *Vs*: el volumen de los datos; la velocidad de recogida y procesamiento en tiempo real; la variedad, formato y taxonomía de los datos, personales o no, obtenidos y tratados procedentes de distintas fuentes; y el valor de tales datos, como los de geolocalización o patrones de conductas.[11]

Pero el tratamiento masivo de datos requiere de tecnologías escalables para el almacenamiento, gestión y análisis eficiente.[12] Y precisamente aquí es donde la implementación de la IA ha adquirido un papel esencial en los últimos años por dos razones fundamentales. En primer lugar, los elementos esenciales que integran la IA son los *datos* y los *algoritmos de aprendizaje automatizado*. En segundo lugar, las distintas técnicas que se engloban bajo la IA no analizan los datos de una manera lineal, sino que aprenden de éstos con la finalidad de inferir determinados modelos a partir de los datos y responder de manera *inteligente* a nuevas entradas de datos adaptando los resultados correspondientes.[13]

2.1 Consideraciones técnicas previas sobre los algoritmos de IA

El Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial (en adelante, por sus siglas en inglés, *AI HLEG*, *High-Level Expert Group on Artificial Intelligence*), creado por la Comisión Europea en 2018, define los *sistemas de inteligencia artificial (IA)* como aquellos

> *programas informáticos (y posiblemente también equipos informáticos) diseñados por seres humanos que, dado un objetivo complejo, actúan en la dimensión física o digital mediante la percepción de su entorno mediante la adquisición de datos, la interpretación de los datos estructurados o no estructurados, el razonamiento sobre el conocimiento o el tratamiento de la información, fruto de estos datos y la decisión de las mejores acciones que se llevarán a cabo para alcanzar el objetivo fijado*.[14]

Por su parte, el reciente borrador de la *Propuesta de Reglamento sobre un Enfoque Europeo para la Inteligencia Artificial* publicado el pasado 21 de abril 2021 en su art. 3.1 define un *sistema IA* como el

> *software que es desarrollado con una o más de las aproximaciones y técnicas enumeradas en el Anexo I y, que para un conjunto de objetivos definidos por un humano, a su vez, puede generar como resultados contenidos, predicciones, recomendaciones, o decisiones que influyen en entornos reales o virtuales*.

Asimismo, entre las técnicas y enfoques de IA que incluye el Anexo I del Borrador de Propuesta de Reglamento se incluyen: (a) Enfoques de aprendizaje automático, incluido el aprendizaje supervisado, no supervisado y por refuerzo, utilizando una amplia variedad de métodos, incluido el aprendizaje profundo; (b) Enfoques basados en la lógica y el conocimiento, incluida la representación del conocimiento, la programación inductiva (lógica), las bases del conocimiento, los motores de inferencia / deductiva, el razonamiento (simbólico) y los sistemas expertos; (c) Enfoques estadísticos, estimación bayesiana, métodos de búsqueda y optimización.[15]

En el caso particular de los algoritmos de ML o aprendizaje automático (subcampo de la IA), estos posibilitan el análisis de los datos para modelar algún aspecto del conocimiento, de manera que las inferencias realizadas a partir de esos modelos se utilizan, a menudo, para predecir y anticipar eventos futuros posibles.[16] A diferencia de la programación tradicional, donde el algoritmo establece las reglas necesarias de una forma determinista para procesar los datos de entrada y obtener unos resultados concretos, en el caso de los modelos de aprendizaje automático el sistema recibe tanto los datos de entrada (*inputs*) como los resultados esperados (*outputs*), a fin de extraer las reglas o lógica que rige la relación entre los *inputs* y los *outputs*, de manera que una vez obtenidas dichas reglas, el modelo las generaliza para aplicarlas a nuevos datos de entrada y producir nuevos resultados. De esta forma, los algoritmos de aprendizaje automatizado son alimentados por un flujo de datos, que contribuyen, con cada iteración, a

identificar el patrón común entre los datos de entrada y los resultados o datos de salida, a partir del cual el algoritmo desarrolla de forma automática el mecanismo de razonamiento. En esencia, esta clase de algoritmos aprende cuál es la estructura de los datos –lo que se conoce como *entrenamiento de los datos*–, y utiliza este aprendizaje para predecir o validar nuevas categorías de datos.[17]

La explotación de los conjuntos de datos con el fin de extraer conocimiento se utiliza para realizar distintos tipos de tareas a fin de resolver problemas diferentes. Estas tareas pueden ser de tipo predictivo (clasificación y regresión), o de tipo descriptivo (*clustering* y asociación). En los problemas de clasificación y regresión existen distintos algoritmos ML que pueden utilizarse y que se mencionan a lo largo de este texto. Salvo en el caso de la regresión logística (*logistic regression*) que sólo funciona con tareas de clasificación, la mayoría de los algoritmos ML más frecuentemente utilizados, como las máquinas de vectores de soporte vectorial (*support vector machine* o SVM), árboles de decisión (*decision trees*), bosques aleatorios (*random forests*), redes neuronales y aprendizaje profundo (*deep learning*), se implementan para resolver tareas de clasificación y regresión. Entre los algoritmos de *clustering* de uso frecuente, está el *K-Means*, que puede implementarse, por ejemplo, para la determinación de la ubicación más óptima de una infraestructura sanitaria, o la elaboración de marcadores en mapas web para agrupar y geoposicionamiento de Bienes de Interés Cultural de una ciudad.[18]

Los dos tipos de aprendizaje más comunes son el supervisado y el no supervisado.[19] En el caso del *aprendizaje supervisado,* el algoritmo (vgr. Naïve Bayes, Radom Forest, K-Means) detecta un patrón o atributo a partir de un conjunto de datos históricos de entrada previamente *etiquetados*, es decir, datos para los cuales se conocen previamente las respuestas o resultados correctos, a fin de predecir el comportamiento de otro conjunto de datos de salida. En el aprendizaje supervisado, cada ejemplo o registro (también denominado *instancia* u *observación*) correspondiente a los datos de entrada debe contener dos elementos: (i) El destino, que es la respuesta que se desea predecir; y las variables, que son los atributos del registro que se utilizan para identificar patrones y predecir la respuesta de destino. Los datos que están etiquetados con el destino (la respuesta correcta) se proporcionan al algoritmo de ML para que identifique y aprenda tales patrones. Y, una vez que el modelo esté entrenado, éste podrá generalizarse para realizar predicciones a partir de nuevos datos para los que el modelo no conoce la respuesta de destino.[20]

En el caso del *aprendizaje no supervisado*, normalmente, mediante técnicas de *clustering* o de asociación, el algoritmo extrae las correlaciones, patrones frecuentes, asociaciones o estructuras existentes en un conjunto de datos sin necesidad de un conocimiento previo sobre qué buscar en los datos. En esta clase de aprendizaje, no es posible establecer *a priori* la relación entre los atributos de entrada y los resultados, sencillamente porque se desconocen esos resultados (los datos no están etiquetados). El aprendizaje no supervisado resulta útil cuando se tratan grandes cantidades de datos no estructurados (e.g. texto, voz, video) y no se tiene un análisis de éstos que permita su clasificación previa. Mientras que, en el aprendizaje supervisado, la máxima sería la *estos son los datos disponibles, ahora muéstrame lo que quiero saber*; en al aprendizaje no supervisado, dicha máxima podría resumirse como *encuentra el patrón existente en estos datos y muéstramelo*.[21]

Por su parte, el *Deep Learning* es un subconjunto de técnicas de ML. Aunque existen distintas técnicas para implementar DL, una de las más comunes es simular un sistema de *redes*

neuronales artificiales que analizan grandes cantidades de datos a través de múltiples capas jerarquizadas de tratamiento de la información, y donde cada capa está entrenada para ser experta en una característica determinada, de manera que cada una pueda reconocer patrones, clasificarlos y categorizarlos y, a partir del ensayo y error, arrojar finalmente una predicción. Una red neuronal simple está compuesta por unas variables de entrada (capa de entrada), una capa de neuronas (capa oculta) donde se van ajustando los pesos (weight) de las conexiones, y las predicciones correspondientes (capa de salida). Por tanto, la red neuronal admite un *input* variable como *información*, un peso variable como *conocimiento* y como resultado una *predicción*. El valor del parámetro *peso* en una red neuronal es una medida de sensibilidad entre los datos de entrada y las predicciones resultantes. Las redes neuronales pueden admitir múltiples datos de entrada al mismo tiempo para generar una predicción. Las redes neuronales se aplican en áreas como el análisis financiero; el diagnóstico médico, el reconocimiento y síntesis de voz; o la clasificación de datos provenientes de sensores.[22]

2.2 ¿Hay *inteligencia* en las Administraciones españolas?

Como podrá intuirse fácilmente, las distintas tareas que pueden realizarse mediante el aprendizaje automático, los tratamientos específicos de datos que tienen lugar durante el entrenamiento del modelo, su modelización mediante diferentes tipos de algoritmos o la determinación del modelo que mejor se ajusta al caso de uso concreto en términos de rendimiento y explicabilidad de las inferencias realizadas pueden tener consecuencias jurídicas de distinto alcance.

En el sector público, estas técnicas de ML tienen un campo de aplicación muy extenso, que van desde la determinación de perfiles de contribuyentes con riesgo de fraude con relación a deducciones fiscales solicitadas;[23] predicción de lugares y momentos de riesgo de comisión de delitos y clasificación de potenciales sujetos criminales en lo que se denomina *policía predictiva*;[24] identificación de establecimientos que deben ser objeto de inspección administrativa, o control de semáforos para optimizar el tráfico de las ciudades; [25] detección y clasificación de incidencias en los servicios públicos de limpieza municipal;[26] procedimientos de contratación y movilidad del profesorado del sistema público de educación[27] y evaluación de su desempeño;[28] gestión de las solicitudes de admisión de alumnos a los estudios universitarios de grado superior;[29] la predicción de riesgo de vulnerabilidad de familias sin hogar a fin de elaborar políticas sociales y asistenciales específicas que faciliten el acceso a una vivienda permanente;[30] o la detección del dolor exagerado y de cardiopatías o el ajuste de la medicación inmunodepresora después de un trasplante en el ámbito de salud pública.[31]

Entre los argumentos a favor de la automatización, total o parcial, de la actividad administrativa, formalizada o no, se dice que la *datificación* y la *algoritmización* pueden facilitar procesos de mejora en la gestión y redefinición de los servicios públicos, contribuir al ahorro de recursos, y dotar de una mayor coherencia al proceso de toma de decisiones, control y evaluación de políticas públicas,[32] así, por ejemplo, las de sostenibilidad y protección ambiental[33] y fomento de la investigación[34] o las de urbanismo, desarrollo de infraestructuras públicas y promoción de vivienda para colectivos vulnerables;[35] o facilitar instrumentos de lucha contra la corrupción y prevención de irregularidades en determinados ámbitos, como en la contratación pública y el fraude fiscal.[36]

Nuestras Administraciones y su sector público institucional, vinculado o dependiente, también se van automatizando y dotando de *inteligencia*, entendida esta última como la capacidad de un sistema de información de adquirir conocimiento oportuno y relevante a partir

de los datos generados localmente u obtenidos externamente gracias a una multitud de fuentes mediante el uso de tecnologías avanzadas que permiten al sistema analizar, predecir y adoptar decisiones de forma total o parcialmente automatizada ante un evento particular, incluso antes de que éste suceda.

No hay más que echar un vistazo a los pliegos de prescripciones técnicas correspondientes a las licitaciones públicas de soluciones tecnológicas avanzadas que se vienen realizando en los últimos años.

Así, por ejemplo, la entidad pública empresarial Red.es licitó en 2015 el desarrollo e implantación de un sistema de Información basado en arquitectura *Big Data* para el análisis y visualización del sentimiento de la población de Castilla La Mancha sobre el SESCAM. La solución debía analizar y visualizar grandes volúmenes de información, estructurada y no estructurada, a partir de reclamaciones, quejas, sugerencias y opiniones que los usuarios del servicio presentaran a través de distintas fuentes (sistemas propios del SESCAM, redes sociales, foros sanitarios y webs de contenido mediático). Entre las funcionalidades a cumplir por la solución, las especificaciones del Pliego de Prescripciones Técnicas incluía: la incorporación de modelos de clasificación de la información; el desarrollo herramientas de detección de conceptos/temas con una aparición más frecuente en un período de tiempo determinado, que permitiera descubrir preocupaciones de los ciudadanos no consideradas de antemano; el desarrollo de modelos de análisis de sentimiento, los métodos de entrenamiento, valoración y reajuste de los modelos, un sistema de *reporting* interactivo para visualizar los resultados; así como la capacidad para, a través de la aplicación de un algoritmo de análisis y clasificación de la información, proporcionar las causas más frecuentes de los tipos extremos de sentimiento.[37]

En el caso de la Agencia Valenciana de Seguridad y Respuesta a las Emergencias, la entidad pública licitó en 2018 el desarrollo y validación de un sistema experto basado en algoritmos ML y DL, de tipo bayesiano y redes neuronales, para clasificar la demanda sanitaria de urgencias, emergencias extrahospitalarias y llamadas al 112, determinando su gravedad, a partir de la información incluida en la base de datos del Sistema de Coordinación de Emergencias y Urgencias Extrahospitalarias (CORDEX).[38]

Por su parte, la Mutua Colaboradora de la Seguridad Social, EGARSAT, licitó en 2020 un contrato que tenía por objeto el diseño, implantación y mantenimiento de sistemas de soporte a la toma de decisiones basados en aprendizaje automático con el objetivo de facilitar la incorporación de servicios y funciones de predicción y análisis avanzado de la duración de bajas por accidente, enfermedad profesional o contingencias comunes, número de visitas a realizar para la optimización de recursos y reducción de la duración de las bajas, así como el número y tipo de bajas segmentadas por diagnóstico o causa y mes. La solución debía incorporar algoritmos de aprendizaje automático para realizar tareas de regresión, *clustering*, clasificación, recomendación-predicción y optimización, a partir de algoritmos de redes neuronales artificiales, bayesianos, bosque aleatorio, o máquina de soporte vectorial (SVM)[39].

Aunque la publicidad de los Pliegos a través de la Plataforma de Contratos del Sector Público o del Perfil del Contratante contribuye a visibilizar el uso de soluciones de aprendizaje automatizado por parte de las Administraciones con diversidad de fines públicos, esta solución no es ni mucho menos satisfactoria. En primer lugar, porque estos instrumentos de publicidad están pensados exclusivamente para garantizar la transparencia y eficiencia del mercado de la

contratación pública, así como la libre concurrencia y no discriminación de los licitadores; y, en segundo lugar, porque estas herramientas no son ni accesibles ni amigables para un ciudadano medio no experto en contratación pública.

Más allá de la farragosa tarea de consultar la Plataforma de Contratación del Sector Público, los Perfiles del Contratante o de alguna iniciativa aislada –como los cincuenta casos documentados de uso de la IA por la Autoridad Catalana de Protección de Datos, en el ámbito de la Administraciones catalanas[40]–, es de lamentar que no exista *un mapeo de los usos de IA en el sector público*.[41] No es una carencia que, en todo caso, pueda ser atribuida en exclusiva a las Administraciones públicas españolas. En el ámbito comparado, también resulta generalizada la a*usencia de una hoja de ruta* que muestre qué sistemas automatizados de IA *están planificando, contratando o desarrollando* las Administraciones.[42]

Pero es que, además, al analizar la contratación de esta clase de soluciones tecnológicas adquiridas por las Administraciones, la doctrina comparada se ha referido a ellas, no sin falta de razón, como proyectos de *llave en mano*, caracterizados por la falta de implicación por parte de la Administración contratante o de un análisis previo del impacto de la solución buscada.[43] A su vez, esta clase de proyectos podría dar lugar a efectos *vendor lock-in* o dependencia tecnológica del contratista, como ha señalado la Oficina de Inteligencia Artificial del Reino Unido en su *Guía para la Compra Pública de Inteligencia Artificial* de 2020; y tal dependencia tecnológica impediría finalmente una adecuada trazabilidad y transparencia de las decisiones adoptadas por la Administración mediante el uso de estos sistemas.[44]

3. El derecho a saber cuánta *inteligencia* utilizan nuestras Administraciones y responsables públicos

A falta de una legislación en nuestro ordenamiento jurídico que prevea alguna limitación, general o particular, respecto de qué potestades (regladas o discrecionales) pueden ejercerse o qué resoluciones pueden adoptarse a través de la IA,[45] o que imponga obligaciones de publicidad activa respecto de la actividad administrativa automatizada,[46] el derecho de acceso a la información pública podría resultar una vía plausible y oportuna –aunque no la única– para conocer qué tipo soluciones tecnológicas avanzadas e inteligentes implementan nuestras Administraciones; hasta qué punto las decisiones públicas que nos afectan las toma una máquina; cómo y a partir de qué datos las toma; cuál es la lógica que subyace detrás de la decisión; si existe o no algún tipo de supervisión humana y, en su caso, hasta dónde llega tal supervisión (*human-in-the-loop*, *human-on-the-loop*, o *human-in-comand*).[47]

3.1 Derecho de acceso, códigos fuente y algoritmos

Según el contexto descrito, cada vez son más frecuente las demandas de ciudadanos y sociedad civil reclamando la publicidad o el acceso al código fuente o a los algoritmos subyacentes implementados por las aplicaciones y sistemas de las Administraciones para la toma de decisiones con claro impacto en *la integridad material e inmaterial de las personas, los grupos y la sociedad en su conjunto*.[48]

Para Coglianese y Lehr, la legitimidad de las elecciones de las Administraciones sobre los algoritmos depende de la transparencia y de la rendición de cuentas. En este sentido, consideran los autores que el principio de transparencia, paso previo y necesario a la rendición de cuentas, es consustancial al Derecho administrativo, y se expresa en normas que, *como la Ley de Libertad*

de Información y la Ley de Procedimiento Administrativo, proporcionan un medio fundamental para que los ciudadanos "sepan" lo que su Administración está haciendo.[49]

A juicio de Desai y Kroll, bajo la lógica de la legislación de transparencia, *ver las partes internas* de un sistema automatizado conduce a la comprensión de su funcionamiento y de las consecuencias asociadas a las operaciones que tienen lugar dentro del sistema. Desde la perspectiva del escrutinio de la actividad pública, un enfoque tal equivale a decir que la forma de obtener evidencias y, por tanto, garantizar que el algoritmo no tiene efectos discriminatorios u otras consecuencias prohibidas o no previstas en la norma implica necesariamente *revisar sus aspectos internos*. De esta posibilidad efectiva de acceder a los aspectos internos del modelo, surge entonces *la capacidad de exigir rendición de cuentas al creador u operador del sistema*.[50]

Descendiendo al ámbito de la casuística, a través del ejercicio del derecho de acceso se han podido conocer determinados *usos inquietantes* de modelos automatizados de toma de decisiones administrativas o de apoyo a la toma de decisiones sobre todo en el ámbito de la llamada *policía predictiva*.

En Estados Unidos, gracias a tres solicitudes de acceso presentadas por la asociación de derechos civiles *Electronic Privacy Information Center* («EPIC»), al amparo de la FOIA, se conocieron detalles relevantes de la implementación por el Departamento de Seguridad Nacional (DHS) norteamericano del Programa *FAST*, *Future Attributes Screening Technology*, que evaluaba atributos psicológicos y comportamentales para determinar la probabilidad de que un individuo, no sospechoso y sin antecedentes penales o policiales, pudiera cometer en un futuro actos delictivos.[51] Ante el silencio de la Administración, en su resolución de medidas cautelares ordenando la puesta a disposición de los documentos que se ajustasen al objeto de la solicitud, resulta elocuente que el Programa en cuestión fuese caracterizado por el Tribunal del Distrito de Columbia como *iniciativa estilo "Minority Report»*, poniendo de relieve la evidente similitud entre el programa FAST y el *Programa Pre-Crimen* de la conocida película de ficción de Spielberg, "Minority Report" (2002).[52]

Una solicitud de acceso ha permitido saber también las aplicaciones concretas en el ámbito de la seguridad pública de un algoritmo de aprendizaje supervisado, como la regresión logística.[53] En el Reino Unido, a resultas de una resolución de estimación parcial del Information Commissioner's Office (ICO) se supo que la Policía de Norfolk utilizaba el mencionado algoritmo para determinar la probabilidad de resolver denuncias por delitos contra la propiedad sobre la base de 29 variables diferentes. Pues bien, de los 971 casos de los delitos evaluados por el algoritmo, el modelo había determinado el archivo de 362 en el periodo comprendido entre enero y septiembre de 2018[54].

Ya en clave nacional, el derecho de acceso nos ha permitido conocer mejor cómo se han automatizado algunos procedimientos sancionadores. Así, a raíz de una reclamación presentada ante el Consejo de Transparencia estatal, en el trámite de alegaciones de las partes, queda documentado todo el procedimiento implementado por el Centro de Tratamiento de Denuncias Automatizadas (CTDA) de la Dirección General de Tráfico (DGT).[55]

Y nuestra jurisdicción contencioso-administrativa aún tiene pendiente de emitir desde mediados de 2019 una resolución desestimatoria del CTBG con relación al acceso al código fuente de la aplicación informática *Bosco.* Dicha aplicación fue puesta en marcha por la

Resolución de 15 de noviembre de 2017 de la Secretaría de Estado de Energía, y permite al comercializador energético de referencia comprobar que el solicitante del bono social puede ser considerado consumidor vulnerable. La cuestión es que la Fundación CIVIO había comprobado que la aplicación excluía sistemáticamente a personas que, con independencia del nivel de renta, cumplían los requisitos legales previstos para ser beneficiarios del bono social: en concreto, las familias numerosas y jubilados con pensiones mínimas.[56] Para comprobar los errores del programa, la Fundación presentó entonces una solicitud de acceso al Ministerio para la Transición Ecológica con el objeto de conocer los siguientes documentos: la especificación técnica de la aplicación; el resultado de las pruebas realizadas para comprobar que la aplicación cumplía con las especificaciones funcionales; el código fuente de la aplicación; así como otros entregables que permitieran conocer el correcto funcionamiento de la aplicación. La solicitud quedó desestimada por silencio administrativo, y ya en trámite de alegaciones ante el Consejo de Transparencia estatal, el Ministerio invocó la inadmisión de la solicitud al considerar que el código fuente no era información pública, al tiempo que aducía la concurrencia de los límites de la seguridad pública y de la protección de los derechos de propiedad industrial e intelectual.[57]

Asimismo, la situación de emergencia sanitaria provocada por el COVID-19 está generando demandas de información de los ciudadanos a sus responsables públicos que tienen que ver con el software o los modelos algorítmicos que se están utilizando con el fin de conocer el impacto real en los derechos y libertades.

A principios de septiembre 2020, la Secretaría de Estado de Digitalización e Inteligencia Artificial (SEDIA) anunciaba la publicación del código fuente de la aplicación móvil para la trazabilidad de contactos, RADAR COVID, con el objeto de iniciar así la *democratización de las infraestructuras digitales públicas*. Más de doscientos miembros de la comunidad académica y científica suscribieron un manifestó en favor de la liberación no sólo del código fuente de la aplicación, sino de la documentación técnica asociada *a fin de que la comunidad científica y la sociedad civil tengan la capacidad de escrutinio necesaria para identificar puntos a mejorar y contribuir a desarrollar y desplegar Radar COVID conforme a los más altos estándares*.[58]

3.2 Matemáticas y programación en la casuística del derecho de acceso

Bourcier y De Filippi nos recuerdan que *el principio de transparencia en la acción administrativa (fundado en el Estado de Derecho) […] se basa en satisfacer este requisito de información, necesaria para que el ciudadano entienda el razonamiento y los datos* que la Administración implementa en sus decisiones; y esto es así, a juicio de los autores, *cualquiera que sea el lenguaje utilizado* por la Administración.[59]

En un contexto hiper-digitalizado y, cada vez más automatizado, la Administración puede expresar sus decisiones y, por tanto, su voluntad también a través del lenguaje de programación insertado en el código fuente de un programa informático.

Al hablar de código fuente y de algoritmos resulta necesario clarificar estos conceptos. Como ha señalado la GAIP, resulta preciso no confundir el algoritmo informático con el código fuente que lo implementa.[60]

En el ciclo de vida completo de un programa, el *código fuente* constituye una de las primeras etapas del material informático que debe producirse; y puede definirse como un conjunto de instrucciones escritas, más o menos en lenguaje natural, siguiendo una gramática, una sintaxis y unas reglas determinadas por un lenguaje de programación,[61] y que, para ser

ejecutadas, deben ser compiladas[62] en el *código objeto*.[63] Por tanto, teniendo en cuenta que el software, generalmente, comprende estas dos formas escritas, tanto el *código fuente* como *el código objeto*» de un programa de ordenador, ambos constituyen formas de expresión de éste.[64]

En el contexto específico de las solicitudes de acceso a la información pública, la CADA ha venido definiendo el *código fuente* como *un conjunto de ficheros informáticos que contienen las instrucciones que deben ser ejecutadas por un microprocesador*, al tiempo que ha calificado tales ficheros informáticos como documentos administrativos comunicables y accesibles para el ciudadano, salvo que concurra algún límite legal aplicable.[65]

Por su parte, para la Autoridad catalana de transparencia, el código fuente sería *el conjunto de instrucciones escritas –en el lenguaje de programación que corresponda– del programa informático empleado para implementar el algoritmo*[66]. Y el CTBG entiende por *código fuente 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*.[67]

En el contexto de litigios en materia de protección de secretos comerciales, los algoritmos han sido definidos, en términos amplios, como una serie de órdenes diseñadas para ejecutar una tarea determinada, o resolver un problema particular. Dado que cada algoritmo establece un orden específico de acciones, algoritmos diferentes pueden conseguir el mismo resultado.[68]

También en el ámbito propio del derecho de acceso a la información pública, la GAIP ha definido los *algoritmos* como el *conjunto finito de reglas que, aplicadas de manera ordenada, permiten la resolución sistemática de un problema, el cual se utiliza como punto de partida en programación informática*.[69]

Las normas administrativas prevén la utilización de algoritmos en muy diversos ámbitos. Así, por ejemplo, cuando el art. 146.2 de la Ley 9/2017, de 8 de noviembre, de Contratos del Sector Público, dispone que, ante pluralidad de criterios de adjudicación, se dará preponderancia a aquellos que hagan referencia a características del objeto del contrato que puedan valorarse automáticamente mediante *la mera aplicación de las fórmulas establecidas en los pliegos*, normalmente, esas fórmulas se identifican en los Pliegos con un algoritmo matemático, cuya corrección y aplicación es fácilmente comprobable por los destinatarios de la resolución de la Mesa de Contratación con la propuesta de adjudicación del contrato.

Sin embargo, el algoritmo implementado por una Administración para la toma de decisiones puede ser mucho más sofisticado y complejo de entender para los propios usuarios del sistema o para los destinatarios de sus decisiones. Y ése el caso de los sistemas de IA que implementan algoritmos de aprendizaje automático.

Desde el año 2017, la policía de Durham en el Reino Unido ha venido desarrollando y testando el sistema HART (*Harm Assessment Risk Tool*) que, a partir de 34 indicadores distintos (la mayoría relacionados con el historial delictivo del sujeto, el código postal e informes policiales), clasifica a los detenidos, según la probabilidad (alta, moderada y baja) de que cometan un delito violento en los dos años siguientes. Sólo aquellos clasificados con perfil de riesgo moderado son remitidos a un programa de rehabilitación, denominado *Checkpoint*, con el fin de descongestionar la jurisdicción penal. Descrito en los términos más sencillos posible, la

herramienta implementa un modelo de aprendizaje supervisado consistente en un *random forest* o de *bosque aleatorio*, construido a partir de 509 árboles de decisión para clasificación y regresión (CART), donde ningún árbol ve todos los datos de entrenamiento completos del conjunto del modelo, sino que cada árbol entrena con distintos registros de datos para elaborar una predicción individual. Cada una de las predicciones recibe un voto, de manera que la predicción con más votos se generaliza y se convierte en la predicción final del conjunto del sistema.[70]

Asimismo, conviene tener presente que la mayoría de los algoritmos ML más frecuentemente utilizados, como el *vecino más cercano* (*Knn nearest neighbour*), *árboles de decisión* (*decision trees*), o *redes bayesianas* (*bayesian networks*) son universales en el sentido de que "pueden aprender cualquier cosa", es decir, pueden aprender a realizar cualquier tarea o función y tomar decisiones, si son entrenados con una cantidad suficiente de datos adecuados. Pero, como explica magistralmente Pedro Domingos, aprender de un conjunto finito de datos implica hacer suposiciones, por lo que diferentes algoritmos ML *hacen diferentes suposiciones, lo que les convierten en buenos para unas cosas y malos para otras.*[71]

La precisión y el rendimiento de los modelos de aprendizaje automatizado dependen, entre otros factores, de la clase de problema a resolver o tarea a ejecutar (e.g. clasificación o regresión, agrupación o *clustering*), de la tipología del conjunto de datos de entrenamiento, del tratamiento previo de esos datos para garantizar su calidad (e.g. completándolos, actualizándolos, eliminando redundancias o datos erróneos, o atribuyendo un valor estimado a aquellos datos que se desconocen), del grado de fiabilidad de la fuente de procedencia de esos datos, del tipo de modelo algorítmico utilizado y su adecuación al caso de uso, así como de los parámetros e hiperparámetros aplicados, o de las métricas utilizadas para evaluar el rendimiento y generalización del modelo para realizar predicciones con relación a nuevos inputs.[72]

Determinar qué modelo de ML y qué componentes del sistema generarán las predicciones más precisas es una cuestión de ejecutar distintas iteraciones sobre los datos y de ajustes del resultado hasta conseguir el menor error.[73] Es más, en la práctica, dependiendo de los problemas a resolver, suelen combinarse distintos modelos ML (*ensemble*), con lo que el resultado, aparte de costoso computacionalmente, puede devenir un modelo de *meta-aprendizaje […] bastante opaco*.[74]

Ahora bien, aunque estos algoritmos de aprendizaje automatizado son los que actualmente suscitan el mayor foco de preocupación por parte de la comunidad internacional, dado su alto impacto en los derechos y libertades individuales y colectivos, Mancosu ha señalado acertadamente que ni legisladores ni Administraciones nacionales están abordando adecuadamente otros algoritmos más elementales, como son los algoritmos deterministas o procedimentales, que se limitan a relacionar unos datos de salida con unos datos de entrada, y que son implementados para una multitud de tareas repetitivas, según unos criterios predeterminados, aunque, en ocasiones, también complejos.[75]

De hecho, el análisis de la casuística jurisprudencial comparada aquí referida sugiere que, en la mayoría de los casos, el foco de la atención del juez no es tanto si el modelo algorítmico impugnado cuyo conocimiento y examen se pretende responde a una arquitectura determinista o de aprendizaje automatizado, cuanto si el interesado ha tenido capacidad de conocer cómo el sistema automatizado o semi-automatizado ha llegado a la decisión atacada y por qué.

En un contexto tal, uno de los principales motivos de preocupación que comporta la generalización de estos modelos automatizados en el ámbito de lo público –en mayor medida, los de IA, pero también los de carácter determinista– es la aparición más que probable de limitaciones relevantes a la *justiciabilidad* de las decisiones administrativas basadas total o parcialmente en los resultados de estos modelos. Como ha sostenido Auby, el juez de lo contencioso no estará mucho mejor capacitado que el ciudadano medio para comprender los complejos procesos de decisión algorítmica y, difícilmente, las técnicas que habitualmente se utilizan para controlar la causa del acto administrativo, su motivación y su racionalidad no siempre resultarán adecuadas para efectuar el control de legalidad de las decisiones basadas en esta clase de sistemas expertos.[76]

4.El código fuente y el algoritmo (determinista o predictivo) utilizados por una Administración son "información pública"

Uno de los argumentos habituales que vienen utilizando las Administraciones para inadmitir solicitudes de acceso al código fuente de una aplicación y/o al modelo algorítmico subyacente es que no constituyen en ningún caso información pública[77] bajo el argumento de que se trata meramente de herramientas de carácter instrumental que se limitan a facilitar el trabajo administrativo interno.[78]

En este sentido, Mancosu ha identificado dos tendencias doctrinales enfrentadas en el ámbito italiano. Por un lado, una mayoritaria que califica los programas informáticos utilizados por la Administración como simples herramientas técnicas, en las cuales los desarrolladores (en la mayoría de los casos, contratista de la Administración) se limitan a ejecutar las instrucciones del poder adjudicador; y por otra, aquella tendencia que califica los programas informáticos en sí mismos como actos administrativos, en la medida en que expresan una voluntad de la Administración condicionada a la ocurrencia del supuesto de hecho previamente identificado y definido en el programa en cuestión.[79]

La distinción entre informática instrumental e informática decisional es relevante porque en el primer caso, no sería procedente el acceso al código fuente al no satisfacer los fines de la legislación de transparencia –el escrutinio del proceso de toma de decisiones de instituciones y responsables públicos,[80] sin perjuicio del derecho a saber si la Administración utiliza una determinada aplicación con fines instrumentales *ad intra* para la realización de su actividad de gestión, si por ejemplo, existen evidencias de que tales aplicaciones pueden plantear eventuales problemas de seguridad y quiebras de la confidencialidad en los sistemas de información de la Administración.

Así, por ejemplo, en el Reino Unido, el ICO estimó una reclamación cuyo objeto era conocer si en la reunión del 31 de marzo de 2020, el *Cabinet Office* mantuvo una reunión virtual utilizando el software para videoconferencia de Zoom, quién había tomado la decisión operativa de utilizar ese software y si realizó alguna evaluación de impacto previa en la privacidad y en la seguridad de la información. Ante la respuesta evasiva del *Cabinet Office*, el *Information Commissioner* señaló que el objeto de la solicitud era muy concreto en términos *objetivos* y que claramente la información solicitada consistía en información pública en el sentido de la FOIA,

es decir, información en poder de las Autoridades públicas *registrada en cualquier soporte*, por lo que apremió al órgano reclamado a contestar en sus justos términos la solicitud.[81]

En el contexto europeo, Francia ha sido uno de los primeros países en reconocer positivamente en su legislación administrativa el derecho de acceso al código fuente de los programas informáticos utilizados por la Administración y, por derivación, el acceso a las reglas que rigen las decisiones adoptadas mediante tratamientos algorítmicos.

En concreto, la llamada *Ley Lemaire* de 2016[82] ha venido a codificar la doctrina del CADA que, previamente, ya había reconocido el carácter de *documento administrativo* del código fuente y de los algoritmos implementados en los programas informáticos utilizados por la Administración.[83] Así, el art. L300-2 del Código de Relaciones entre el Público y la Administración (CRPA) califica el código fuente utilizado por una Administración como *documento administrativo*.

En coherencia con lo anterior, la CADA ha calificado como *documentos administrativos* no sólo el código fuente[84] o los algoritmos implementados por una Administración,[85] sino también la documentación técnica relativa al código fuente, como puede ser el documento de especificación de requisitos de software.[86]

Así, por ejemplo, la Autoridad francesa ha estimado el derecho de acceso al código fuente de la plataforma Parcoursup habilitada para la gestión automatizada del procedimiento nacional de preinscripción en el primer año de enseñanza universitaria pública, junto con las especificaciones del software, *presentadas sintéticamente*, y el algoritmo de procesamiento.[87]

Asimismo, el código fuente del programa utilizado por el Fondo Nacional de Subsidio Familiar (CNAF) para el cálculo completo de prestaciones legales (ayudas financieras de carácter familiar o social), administradas por la red de cajas locales de asignación familiar, ha sido calificado también por la CADA como *documento administrativo*, junto con los archivos SQL de dicho código fuente y las especificaciones funcionales utilizadas para calcular las ayudas a la vivienda, la renta solidaria activa, las asignaciones familiares y a la prima de actividad.[88]

La Autoridad catalana, con una línea interpretativa muy abierta, ha considerado que el concepto de información pública *trasciende el tradicional de documentos y es sustancialmente equivalente al de conocimiento*, de manera que el derecho de acceso no sólo se proyecta sobre los documentos en poder de la Administración, sino también sobre otros soportes de conocimiento también detentados por la Administración como pueden ser *las bases de datos informáticas, los algoritmos o conocimiento material no formalizado en documento alguno o registro determinado*.[89]

En consecuencia, la GAIP ha estimado que tanto el código fuente como el algoritmo implementado por un programa informático empleado por la Administración constituyen información pública a los efectos tanto del artículo 2.b de la Ley 19/2014, del 29 de diciembre, de transparencia, acceso a la información pública y buen gobierno de Cataluña (LTAIBG-Cat) como del art. 13 LTAIBG estatal. Para la Autoridad catalana, ambos preceptos legales incluyen en el concepto de *información pública* toda aquella información que la Administración elabore o tenga en su poder en ejercicio de sus funciones, con independencia del lenguaje o forma en que se exprese, lo que incluye, entre otros, el lenguaje matemático o el informático.[90]

5. Riesgos inherentes a la actuación administrativa automatizada. El interés público en el derecho de acceso

A pesar de las bondades predicadas de la *datificación* y la a*lgoritmización* de la actividad administrativa, los sistemas automatizados, total o parcialmente, de toma de decisiones –especialmente los basados en algoritmos ML, implican una serie de riesgos inherentes para los derechos y libertades de los ciudadanos.

Entre las críticas más habituales a esta clase de sistemas automatizados basados en la IA se considera, en primer lugar, que estos sistemas pueden favorecer la reglamentación oculta, *extra legem y contra legem*, como consecuencia de la traducción incorrecta, querida o no, de la norma jurídica en lenguaje de código. Sin perjuicio de su eficacia, *la formalización y la transposición de las normas jurídicas a un lenguaje informático puede afectar no sólo a la naturaleza, sino también al alcance y la significación de estas normas*, en la medida en que el diseñador del modelo, normalmente, el experto informático, ajeno por otro lado a cualquier control democrático, al traducir la norma en código, incorpora inevitablemente su propia interpretación subjetiva de la norma, con el riesgo de malinterpretar el significado de la misma, diseñando, así, involuntaria o intencionadamente, un modelo que modifica el contenido y alcance de la norma.[91]

En segundo lugar, la automatización de decisiones puede socavar las garantías del interesado en el procedimiento administrativo, reforzar los sesgos discriminatorios y desigualdades ya existentes e, incluso, distorsionar la finalidad de las políticas públicas,[92] como consecuencia, por ejemplo, de un modelo desarrollado a partir de unos datos de entrada o entrenamiento incompletos o no representativos, o la selección de datos procedentes de fuentes sesgadas.[93]

En tercer lugar, otros efectos adversos de estos modelos automatizados inteligentes serían la denominada *automatización del sesgo*, esto es, la complacencia y dependencia excesiva del usuario del modelo (en este caso, la Administración) con los resultados generados por el algoritmo sin entrar a evaluar la adecuación, validez y justicia del modelo y sus resultados, así como la falta de interpretabilidad de las decisiones adoptadas total o parcialmente mediante estos modelos[94], lo que podría limitar seriamente el derecho del interesado a impugnar la decisión[95] y, por ende, el control jurisdiccional de las decisiones administrativas basadas en los resultados de estos sistemas.[96]

5.1. Reglamentación oculta y otros *bugs*

El primer y más inmediato riesgo inherente a la actuación administrativa automatizada se produciría con la traducción de la norma jurídica en lenguaje de código, pues esta operación implica que el diseñador de un sistema automatizado, ya sea de soporte a las decisiones administrativas o de adopción autónoma de tales decisiones, *debe analizar el discurso legal para representar no solo la información (las proposiciones) sino el razonamiento (el motor de inferencia) que implica el texto (legal)*"[97]

Al proceso mediante el cual *las fuentes del Derecho expresadas en lenguaje natural son remodeladas en representaciones de la norma mediante el lenguaje de programación* se ha referido Schartum como *transformación de fuentes*. En la medida en que los lenguajes de

programación son precisos e inequívocos, la transformación implica que las incertidumbres y la flexibilidad en la interpretación de las normas deben ser sustituidas por un conjunto de reglas legibles por máquina, precisas y fijas. Lo anterior significa, por un lado, que los sistemas automatizados de decisiones administrativas son, a menudo,

> *el resultado de super-complejos procesos de interpretación y representación de la normativa relevante, y es esta representación algorítmica -y no la auténtica norma- la que es aplicada a los casos individuales y la que decide los resultados [subrayado nuestro].*

Por otro, las operaciones tradicionales de interpretación, integración y aplicación de las normas jurídicas se vuelven mucho más opacas e ininteligibles en este proceso de transformación.[98]

Parece lógico pensar que esta representación algorítmica de las fuentes del Derecho resultará especialmente compleja cuando se trate de transformar en el rígido y formalizado lenguaje de programación los conceptos jurídicos indeterminados, los preceptos en blanco (que deben ser integrados por la norma de remisión), las lagunas legales, o el ejercicio de potestades discrecionales contenidas expresa o implícitamente en las distintas fuentes del Derecho, puesto que el significado de esos conceptos, la integración de la norma de remisión o los márgenes de apreciación de las posibles soluciones igualmente justas raramente aparecen como tales en la literalidad de la norma, y son fruto de una operación «intelectual/humana» de ponderación de intereses en conflicto. Pero, incluso, en el caso de los actos reglados, pueden existir elementos que exigen la interpretación e integración de la norma aplicable.

Como ha sugerido De la Quadra-Salcedo, con relación a las Administraciones públicas resulta especialmente relevante considerar *la eventual existencia de errores en el modelo base para la construcción del algoritmo o de los datos que se toman en cuenta en relación con su inadecuación al fin que se pretende o la licitud del fin*.[99] Aunque esta aproximación nos sitúa en el plano jurídico de los distintos grados de invalidez del acto administrativo, no cabe duda de que los aspectos técnicos relativos al diseño, configuración e implementación de un modelo ML, ya sea para la toma de decisiones de forma automatizada o como soporte a la toma decisiones en el ámbito de la actividad administrativa, resultan necesariamente determinantes.

En algunos asuntos abordados por la casuística judicial norteamericana se ponen ya de manifiesto algunas de las implicaciones jurídicas que pueden tener los errores técnicos en el diseño e implementación de estos modelos en el ámbito público. En *K.W. v. Armstrong* (2016), el Tribunal de Distrito de Idaho consideró que la implementación por el Departamento Estatal de Salud y Bienestar del Estado (IDHW) de una herramienta informática para solicitar las prestaciones para personas con discapacidad física o mental correspondientes al sistema de ayudas públicas, *Medicaid*, vulneraba el derecho de los interesados al procedimiento legalmente establecido al reducir de forma arbitraria e injustificada la cuantía de la prestación, impidiendo la impugnación efectiva de las resoluciones administrativas que asignaban la cuantía de la prestación final. Aunque a lo largo del procedimiento judicial no se llegó a identificar el modelo algorítmico utilizado por la herramienta para el cálculo de las prestaciones, para el Tribunal quedó absolutamente acreditada la *falta de fiabilidad* de la herramienta por la existencia de *errores considerables de origen desconocido* y la *ausencia de un control de calidad* de la herramienta. Asimismo, se consideró que el procedimiento de impugnación de las resoluciones asignando la cuantía de las prestaciones mediante la herramienta informática, *lejos de ser robusto* resultaba extremadamente complejo para los solicitantes.[100]

El planteamiento anterior, sin duda, nos sitúa directamente ante lo que algunos han denominado como la *desviación informática de poder*; y, en términos más amplios, ante el problema del control de vicios de nulidad o anulabilidad de los actos administrativos, la articulación del principio de restricción de la invalidez o la existencia de irregularidades no invalidantes en la decisión administrativa automatizada.[101]

A lo anterior deben añadirse los riesgos adicionales derivados de la licitación frecuente de esta clase de soluciones tecnológicas por parte de la Administración, al incorporar modelos y *know how* desarrollados por el contratista puede estar sujeto a derechos exclusivos de propiedad intelectual o industrial o al secreto comercial, dado que habitualmente *los responsables públicos y funcionarios carecen del necesario entendimiento de cómo funcionan esos sistemas y cómo mitigar los daños causados por los errores o sesgos embebidos en el sistema*"[102]

A la existencia de errores o *bugs* embebidos en los modelos algorítmicos y a sus consecuencias en los derechos del interesado se han empezado a referir también algunos Tribunales nacionales en el ámbito europeo.

Situados ahora en el ámbito doméstico, el planteamiento de nuestros Consejos de Transparencia sobre esta cuestión de los errores técnicos y su incidencia en los derechos de los administrados ha resultado desigual, como así lo ha sido también la determinación del interés público existente en el ejercicio del derecho de acceso en estos casos.

Con relación a los tratamientos automatizados de detección de infracciones por exceso de velocidad por el CTDA y generación de los expedientes sancionadores en la DGT, comentado *supra*, una de las cuestiones que se plantean en el trámite de audiencia al interesado tiene que ver con las consecuencias de los errores embebidos (*bugs*) en los sistemas automatizados de toma de decisiones y su influencia en la resolución sancionadora. Cuestión, sin embargo, desestimada por el CTBG por entender que la reclamación especial ante la Autoridad de control no es el cauce adecuado para solicitar el acceso a dicha información.[103]

El planteamiento de la Autoridad catalana dista del anterior, al estimar el derecho de acceso tanto al código fuente como al algoritmo subyacente de un programa informático empleado por el Consejo Interuniversitario de Cataluña para la designación de los miembros de los Tribunales evaluadores de las pruebas de acceso a la universidad (PAU). Entiende la GAIP que

> *existe un interés público y privado –de los interesados que participan en las diversas convocatorias– evidente en poder comprobar que el programa informático está correctamente diseñado para garantizar la igualdad de todos los participantes y que la designación de los miembros de los tribunales se ajusta a los criterios establecidos por la normativa que los regula.*

Tal interés público y privado se concreta en verificar si el código fuente se ha limitado *a recoger y aplicar correctamente las variables antes mencionadas (requerimientos de paridad entre hombres y mujeres y de porcentajes mínimos de profesores universitarios y de bachillerato,*

etc.), de forma reglada, de acuerdo con la norma jurídica específica que regula el procedimiento de selección de los miembros de los tribunales correctores. El interés público que justificaría el derecho de acceso se ve reforzado, además, por el hecho de que los miembros de los tribunales correctores de las pruebas de acceso a la Universidad ejercen una función pública relevante y son retribuidos con dinero público.[104]

En el caso del bono social, aun estimando el acceso parcial –limitado fundamentalmente a las especificaciones técnicas y a las pruebas realizadas para verificación de funcionalidades–, sin embargo, el CTBG no entra a justificar el interés público existente en el derecho de acceso, más allá de las finalidades generales de la Ley de Transparencia.[105] Posiblemente esto sea así, porque en realidad el Consejo rechaza de plano la concurrencia de los límites alegados por el Ministerio para la Transición Ecológica. Ya en sede contencioso-administrativa, CIVIO explica cuáles son los riesgos inherentes a los errores embebidos en los sistemas automatizados implementados por la Administración:

> *En el caso que nos ocupa, a la Fundación Ciudadana Civio se le ha hecho entrega de las funcionalidades del programa solicitado, que es a través del cual se decide si se tiene derecho al bono social de la energía eléctrica. Y, lo que se evidencia de lo entregado es que el programa comete un error. Tal y como consta en el Anexo II, los posibles resultados de la aplicación son: concedido (vulnerable, vulnerable severo), no cumple [con los requisitos] y en algunos casos da como respuesta «Imposibilidad de cálculo»: esto se debe a que, tal y como está diseñada la aplicación, siempre exige comprobar el nivel de renta de la unidad familiar y cuando no puede comprobarlo devuelve la respuesta de «Imposibilidad de cálculo». El problema es que hay casos en los que el solicitante tiene derecho al bono social independientemente de su nivel de renta (por ejemplo, jubilados con pensiones mínimas o familias numerosas) y la aplicación en vez de otorgarles el bono les responde con "Imposibilidad de cálculo".*

Detrás del planteamiento de CIVIO subyace, sin duda, la concepción del código fuente y de los algoritmos que lo implementan como auténticas normas jurídicas: *[…] cuando el código fuente de un programa informático es ley, porque mediante su ejecución se generan derechos y obligaciones, el ciudadano tiene tanto derecho a inspeccionar su funcionamiento como lo tiene con respecto a cualquier otra norma jurídica.* El fundamento de este derecho a inspeccionar el código fuente serían los principios de legalidad, publicidad de las normas e interdicción de la arbitrariedad. *El código fuente, en este aspecto, ha de ser sometido a los mismos requisitos –señala CIVIO en su recurso– que los que se imponen a las fuentes del derecho*.[106]

Dejando al margen la discutible naturaleza reglamentaria o legal del código fuente y los algoritmos que lo implementan, en los planteamientos anteriores (de la GAIP o del recurso de Civio) subyace la idea de que el acceso al código fuente de un programa informático empleado por la Administración posibilitaría comprobar si dicho programa está correctamente diseñado o no y, por ende, si los parámetros funcionales del programa en cuestión –traducidos a un lenguaje de programación concreto– cumplen o no con las finalidades previstas en la norma jurídica que implementa dicho programa.

5.2. La *vis* expansiva del *black box* decisional

Dado que en los sistemas ML, y particularmente, en los de DL, se necesitan un conjunto de datos suficientemente grande y representativo del problema a resolver para desarrollar su propio sistema de razonamiento, no resulta siempre fácil entender las razones de las decisiones tomadas por el modelo algorítmico y, en consecuencia, garantizar su inteligilidad.[107]

En este sentido, Desai y Kroll señalan que, en buena parte de la literatura científica existente la crítica habitual es que el ciudadano no puede entender las decisiones adoptadas mediante sistemas automatizados que implementan algoritmos IA, precisamente porque *el proceso de toma de decisiones es un black box*[108]. El problema de la "Caja Negra" puede ser definido como *la incapacidad para comprender totalmente un proceso de toma de decisiones mediante IA y la incapacidad de predecir las decisiones o resultados de un sistema de IA*, incluso para el experto humano que diseñó el sistema.[109]

En el Reino Unido, el *Information Commissioner Office* y el Instituto Alan Turing han elaborado una interesante clasificación de los algoritmos de aprendizaje automatizado según su nivel de explicabilidad u opacidad. Así, por ejemplo, los algoritmos bayesianos, de regresión logística o de árbol de decisión se caracterizarían por su *buen nivel de interpretabilidad*. En cambio, junto con el ensamblado de modelos, más arriba explicado, la máquina de soporte vectorial (SVM) y el *random forest* presentarían niveles muy bajos de interpretabilidad. Y, finalmente, en el caso de las redes neuronales, éstas se califican como el *epítome de las técnicas de blackbox*[110], pues aprenden las relaciones existentes entre datos y patrones a través de su estructura en capas desarrollando sus propias reglas decisionales habitualmente ininteligibles para los humanos.

En el caso de las redes neuronales, Liu *et al.* han explicado que, si bien el algoritmo reduce el esfuerzo computacional a realizar por el sistema experto, aumentando la precisión en la extracción y análisis de patrones procedentes de conjuntos masivos de datos, lo hace a costa de la *capacidad humana de explicar cualitativamente el razonamiento inferencial que sucede en cada nivel neuronal*.[111]

Por su parte, el Grupo de Expertos Independientes de Alto Nivel sobre IA de la Comisión Europea describe de una manera muy elocuente los problemas de interpretabilidad que plantean las redes neuronales:

> *Los procesos de formación con redes neuronales pueden dar lugar a parámetros de red configurados con valores numéricos difíciles de correlacionar con los resultados. Además, a veces unas pequeñas variaciones en los valores de los datos pueden traducirse en interpretaciones completamente diferentes, provocando, por ejemplo, que el sistema confunda un autobús escolar con un avestruz*.[112]

Como ha puesto de relieve De Laat, en la práctica, la relación entre precisión e interpretabilidad no siempre es pacífica en determinados modelos de *Machine Learning*, pues aquellos que alcanzan mayor precisión lo hacen a costa de relegar la interpretabilidad de los resultados a un segundo plano. Así, por ejemplo, para realizar tareas de clasificación no se utiliza

un simple árbol de decisión, sino el sumatorio de hasta cientos de ellos, de manera que no es posible determinar de manera directa cuál de los árboles del bosque aleatorio ha contribuido en mayor medida a generar un determinado resultado. En el caso de las redes neuronales, los pesos que conectan las variables de entrada a las variables intermedias, así como los pesos que conectan las variables intermedias con los resultados son ajustados mediante diversas iteraciones. Sin embargo, aunque el modelo final muestra todos esos pesos, no es posible interpretar en qué medida las distintas variables de entrada contribuyen a un determinado resultado final.[113]

Si traducimos al lenguaje jurídico-administrativo todo lo anterior, puede decirse que a mayor falta de *interpretabilidad técnica*, mayor riesgo de que la decisión administrativa pueda incurrir en algún supuesto de invalidez. A lo que debe añadirse la imposibilidad del interesado de verificar la corrección y legalidad de la decisión y, en última instancia, atacar los eventuales vicios que afectan a la validez y su derecho a la tutela judicial efectiva.

Pero es que la falta de interpretabilidad de los modelos algorítmicos de aprendizaje automatizado está ahí, con independencia de que el modelo implementado sea o no de *caja negra*. De hecho, un modelo altamente interpretable, como el *árbol de decisión*, puede perder su nivel de *explicabilidad local* cuando el modelo tiene una alta dimensionalidad al incorporar muchas variables y relaciones imposibles de identificar y verificar para el razonamiento humano. En estos casos, el algoritmo puede tornarse extremadamente opaco, si no se facilita la suficiente información sobre el procedimiento de testado y validación del modelo acompañado de herramientas específicas complementarias que proporcionen una explicación comprensible.[114]

Las consecuencias de una actividad administrativa algorítmica, opaca e incomprensible, para los destinatarios de las decisiones administrativas empiezan a ser evaluadas por los Tribunales, con independencia de que la decisión sea total o parcialmente automatizada o de que el algoritmo implementado sea determinista o predictivo. Más aún, el concepto técnico de *black box* generado en el contexto de la IA empieza a aplicarse no sólo para los algoritmos ML o DL, sino a cualquier modelo total o parcialmente automatizado de toma de decisiones, al margen del tipo de algoritmo implementado, cuando no es posible verificar la corrección y la adecuación a Derecho de las decisiones adoptadas por el modelo.

El Consejo de Estado de los Países Bajos se pronunció, *obiter dicta*, y por primera vez, sobre un procedimiento semi-automatizado de concesión de autorizaciones administrativas en materia ambiental, donde las predicciones del algoritmo se utilizaban como apoyo a la decisión estimatoria o desestimatoria de la autorización. El Consejo de Estado consideró que el uso del software en cuestión implicaba un claro riesgo de falta de transparencia y verificabilidad del procedimiento parcialmente automatizado en que se basaban las decisiones administrativas, *debido a la falta de conocimiento de las elecciones realizadas, así como los datos y parámetros utilizados*. Argumentaba el Consejo que, si las partes interesadas deseaban impugnar las decisiones adoptadas sobre la base de ese procedimiento parcialmente automatizado, se encontraban en una clara *situación de desigualdad de armas*, pues el programa informático utilizado podía ser considerado *como un black box desde la perspectiva de quien pretende atacar la decisión*. Con el fin de impedir esta situación de desigualdad procedimental, el Consejo concluía que el Estado estaba obligado a hacer accesible a terceros la información relativa a las elecciones, datos y parámetros de una manera completa, oportuna y adecuada [subrayado nuestro]"; y, en su caso, motivar la decisión adoptada, posibilitando así un *amparo legal genuino frente a decisiones basadas en esas elecciones, datos y parámetros*, y el debido control judicial de la decisión.[115]

5.3. Sesgos embebidos y vulneración de derechos

Otra de las críticas habituales a los modelos algorítmicos de IA suele ser la cuestión del sesgo. Los sesgos pueden ser *preexistentes*, al reflejar percepciones y prácticas sociales discriminatorias; pueden ser *técnicos*, es decir, atribuibles a las limitaciones de los sistemas informáticos; o *emergentes*, y en tal caso, sólo son detectables después de que los usuarios hayan interactuado con el algoritmo. En consecuencia, hacer los algoritmos transparentes para un amplio rango de posibles interesados, y en particular, para aquellos afectados por los resultados del modelo implementado es crucial para identificar y gestionar los sesgos.[116]
Desde la perspectiva de la creciente *algoritmización de la Administración pública*[117] no puede obviarse, además, que el riesgo de sesgo puede darse en cada una de las *decisiones atomizadas* que toman los algoritmos de aprendizaje automático.[118]

En estos modelos de aprendizaje automatizado, donde las reglas que rigen el código son inherentemente dinámicas y adaptativas, la evolución autónoma de tales modelos, redefiniendo constante y autónomamente las reglas a partir de los inputs recibidos, customizando y adaptando el perfil de los destinatarios individuales de la toma de decisiones automatizada, puede quebrar el principio básico de igualdad ante ley y de no discriminación,[119] y en consecuencia, producir efectos *contra legem* al limitar o impedir el ejercicio de derechos que están amparados en la norma jurídica.[120]

A título ejemplificativo que no limitativo, el impacto de modelos automatizados implementados por el sector público en los derechos de los ciudadanos en general, y en los derechos y garantías de los interesados en un procedimiento administrativo, en particular, queda evidenciado en ámbitos distintos de la actividad administrativa.

Como hemos visto más arriba, los modelos de ML más eficientes o la combinación de estos (*ensemble*) pueden resultar muy opacos. Así, por ejemplo, los sistemas automatizados de reconocimiento facial, cada vez más utilizados por Fuerzas y Cuerpos de Seguridad con fines de investigación y prevención de delitos o para el control de fronteras e inmigración, son foco de preocupación creciente, pues implican graves riesgos no sólo para la privacidad, sino también para otros derechos y libertades de los ciudadanos (la libertad, la seguridad, la libertad de expresión, el derecho de manifestación), dado que, en la práctica, se producen falsos positivos o falsos negativos.[121]

Así las cosas, el pasado 1 de enero de 2021 entró en vigor una Ordenanza local en la Ciudad de Portland (Oregón) que prohíbe expresamente la adquisición y uso de tecnologías de reconocimiento facial por las autoridades locales. En concreto, la Ordenanza 190113 que prohíbe tales usos a las Autoridades locales señala que:

> *El uso de las Tecnologías de Reconocimiento Facial plantean preocupación en torno a la privacidad, intrusión y falta de transparencia. La falta de transparencia y rendición de cuentas, junto con tecnologías sesgadas –particularmente en el contexto de los falsos positivos con relación a su uso con fines de investigación policial– puede generar impactos devastadores en individuos y familias*.[122]

El caso *Williams* es un claro ejemplo de las consecuencias personales de un falso positivo, cuando la Policía de Detroit detuvo erróneamente en enero de 2020 a un ciudadano afroamericano porque un algoritmo de reconocimiento facial había detectado una posible coincidencia entre las imágenes procedentes de una cámara de videovigilancia –donde aparecía un sujeto atracando una tienda dos años antes– y la imagen fotográfica del permiso de conducir del afectado obrante en la base de datos de la Policía estatal.[123] La gran ironía es que, mucho antes del incidente, en septiembre de 2019, varias organizaciones de derechos civiles habían presentado una solicitud de acceso al amparo de la *Freedom Information Act* de 1966 a las autoridades de Detroit. Tal solicitud tenía por objeto conocer la información relativa al contrato de adquisición de un software de reconocimiento facial en tiempo real por videovigilancia entre la Policía local y la empresa contratista desarrolladora.[124]

También, al amparo de la FOIA, la *American Civil Liberties Union Foundation* («ACLU») solicitó información a una serie de agencias federales (entre otras, al Departamento de Justicia, al FBI, al Departamento de Seguridad Interior, a la Oficina de Aduanas y Protección Fronteriza, o al Servicio de Ciudadanía e Inmigración) con relación a la monitorización de inmigrantes y solicitantes de visados en redes sociales, así como al intercambio de esta información entre las agencias federales reclamadas. En concreto, ACLU solicitó el acceso a información relativa a la adquisición de tecnologías de vigilancia de redes sociales y, particularmente, al *uso o agregación de contenidos de redes sociales en sistemas o programas que utilicen algoritmos, tratamientos machine-learning o aplicaciones de análisis predictivo* [subrayado nuestro]. Invocando la llamada *doctrina Gloma,r*[125] basada en la excepción (7)(E) FOIA,[126] el FBI ni confirmó ni negó la existencia de información objeto de la solicitud realizada por ACLU. Sin embargo, para el Tribunal del Distrito Norte de California, la Agencia federal no justificó adecuadamente la procedencia de esta doctrina.[127]

6. Límites al acceso al código fuente y a los algoritmos públicos

De la casuística comparada e interna existente sobre el ejercicio de derecho de acceso al código fuente y algoritmos de los programas implementados por la Administración, entre las causas de inadmisión y límites habitualmente invocados –no siempre correctamente[128]– para denegar el acceso deben destacarse, entre otros, la necesidad de acción previa de reelaboración; el coste de la tramitación de la solicitud; la seguridad pública; la prevención, investigación y sanción de ilícitos; la protección de los intereses comerciales o el derecho de propiedad intelectual.

Entre lo que la legislación de transparencia española suele denominar *causas de inadmisión*, la casuística comparada ofrece algunos ejemplos significativos.

Así, por ejemplo, según el Tribunal de Distrito de Columbia, no sería objeto de acceso la información relativa al resultado de un *algoritmo confidencial* utilizado por la Administración Federal de Aviación para identificar el número de registro de una aeronave a partir del código del transpondedor del aparato, pues la *FOIA no impone un deber a la agencia de elaborar información*.[129]

En el asunto del algoritmo de solubilidad implementado por la Policía de Norfolk se consideró que la tramitación de la solicitud excedía del coste límite apropiado,[130] por lo que se denegó el acceso a la información relativa al número de casos archivados por el algoritmo en los que el funcionario responsable de su revisión, sin embargo, habría determinado el seguimiento de la denuncia, y de entre estos últimos, el número de sospechosos que, finalmente, habrían

sido detenidos/interrogados; así como la información relativa al listado de los factores de riesgo aplicados por el algoritmo de solubilidad y los informes sobre la ejecución del algoritmo. En particular, el ICO calculó que facilitar tal información en los términos requeridos por el solicitante conllevaría 33,8 horas de tramitación de la solicitud, lo que excedía del límite reglamentario establecido.[131]

6.1. El límite de la seguridad pública

Uno de los límites frecuentemente invocados frente al acceso al código fuente, y en su caso, al algoritmo subyacente, es la *seguridad pública,*[132] incluiría también el concepto de seguridad lógica o informática.

Entrando en la casuística específica comparada, la CADA ha desestimado el acceso al código fuente de *SAIP* –Sistema de Alerta y de Información de la Población– del Ministerio del Interior francés, una aplicación móvil para smartphones para la notificación de mensajes de alertas ante la sospecha de eventuales atentados o eventos excepcionales de seguridad civil, así como consejos e instrucciones a seguir por la población en tales casos. [133] La Comisión consideró que la divulgación del código fuente solicitado facilitaría ataques contra dicha aplicación y consiguiente neutralización en una situación de atentado, poniendo en peligro la seguridad pública y la seguridad de las personas.[134]

En sentido similar, también ha desestimado el acceso al código fuente de la aplicación ALICEM, desarrollada por el Ministerio del Interior y la Agencia Nacional de Títulos Securizados (ANTS) que permite el acceso del ciudadano a los servicios públicos en línea disponibles en la plataforma FranceConnect mediante la autenticación certificada de la identidad a través del reconocimiento facial biométrico y tecnología NFC (Near Field Communication). La Comisión de Acceso considera que, en este contexto, el código fuente de *ALICEM* incorpora procedimientos que permiten, durante diferentes fases, garantizar un alto nivel de seguridad de la aplicación y contribuir a la lucha contra el fraude documental y la usurpación del estado civil. Por tanto, el código fuente en sí mismo, constituye uno de los "*factores de seguridad de la aplicación informática*", a lo que debe añadirse que dicho código está sujeto a una "*distribución restringida en el sentido del Decreto de 23 de julio de 2010, por el que se aprueba la Instrucción General interministerial sobre la Protección de los Secretos de la Defensa Nacional*". En vista de lo anterior, la CADA finalmente desestima el acceso, al entender que *la divulgación del código fuente solicitado probablemente debilitaría la seguridad de la aplicación "ALICEM" y haría más vulnerables a sus usuarios*.[135]

En nuestro ordenamiento jurídico, la *seguridad pública* constituiría un título de intervención en el ámbito de la protección de las infraestructuras informáticas con que cuentan las Administraciones para el ejercicio de sus potestades, por lo que el acceso a esta clase de información no sólo podría suponer un riesgo para la seguridad lógica de las redes y sistemas de información de las Administraciones,[136] sino también de *gaming* o engaño a los sistemas basados en datos para eludir la aplicación de la Ley.[137]

En este sentido, el Esquema Nacional de Seguridad (ENS) define la *seguridad de las redes y de la información* como

> *la capacidad de las redes o de los sistemas de información de resistir, con un determinado nivel de confianza, los accidentes o acciones ilícitas o malintencionadas que comprometan la disponibilidad, autenticidad, integridad y confidencialidad de los datos almacenados o transmitidos y de los servicios que dichas redes y sistemas ofrecen o hacen accesibles*.[138]

Teniendo en cuenta que las aplicaciones utilizadas por la Administración son activos de sus sistemas de información susceptibles de ser atacados deliberada o accidentalmente con consecuencias potencialmente adversas o críticas, de acuerdo con el ENS,[139] en una eventual ponderación entre el derecho de acceso y el límite de la seguridad pública deberían tenerse en cuenta las dimensiones de seguridad (disponibilidad, autenticidad, integridad, confidencialidad, trazabilidad) que podrían verse comprometidas según la categoría del sistema de información (Alto, Medio, Bajo) afectado por el acceso al software en cuestión, las medidas de seguridad exigidas y la repercusión en la capacidad de la Administración para el logro de sus objetivos, la protección del activo en cuestión y de otros activos en función de su dependencia, el cumplimiento de sus obligaciones de servicio, así como el respeto de la legalidad y los derechos de los ciudadanos.[140]

Las consideraciones previas son muy pertinentes, pues con distinto alcance, por las eventuales implicaciones en la seguridad de los sistemas de información de las Administraciones, se han pronunciado la GAIP y el CTBG.

A la hora de determinar si el acceso por un tercero al código fuente del programa informático implementado por el Consejo Interuniversitario de Cataluña (CIC) en la designación de los miembros de los Tribunales correctores de las Pruebas de Acceso a la Universidad (PAU) podría afectar la *seguridad pública*, la Autoridad Catalana estima que debe tomarse en consideración *la naturaleza y la finalidad del programa informático solicitado* con relación al posible daño o perjuicio que la divulgación del código fuente pudiera comportar. En su respuesta estimatoria en favor del acceso, la GAIP considera que,

> *[s]i de lo que se trata es evitar que el código fuente pueda ser manipulado por terceros, esto no se garantiza impidiendo su conocimiento, sino adoptando las medidas de seguridad necesarias para evitar que terceras personas puedan acceder –presencialmente o de forma remota– a los ordenadores y sistemas informáticos que lo utilizan*.

En la resolución favorable al acceso, resulta determinante que, durante la tramitación de la reclamación, el *Centre de Seguretat de la Informació de Catalunya* (CESICAT) –actual *Agència de Ciberseguretat de Catalunya*– no respondiera a la solicitud del Consejo Interuniversitario de evacuación del informe correspondiente para que, en su condición de organismo especializado, el CESICAT se pronunciara sobre los eventuales riesgos que, para la seguridad pública, pudiera ocasionar la estimación del derecho de acceso al código fuente.[141]

Asimismo, con relación a la solicitud de la Fundación CIVIO de acceso al código fuente y a las especificaciones técnicas de la aplicación informática del bono social, el CTBG se limitó a subrayar, sin entrar en el fondo del asunto, que resultaba admisible la invocación *genérica* de los límites de la *defensa nacional* y de la *seguridad pública* por parte de la Administración, *sin argumentar mínimamente porqué resultan de aplicación a su juicio*.[142] Respecto del límite de la seguridad pública, el Consejo consideró insuficiente el siguiente argumento desestimatorio del Ministerio para la Transición Ecológica:

> *El acceso a esta documentación supondría facilitar a un tercero determinada información que afectaría a la seguridad de la propia Administración, ya que a través del código fuente y de las propias especificaciones técnicas se dan detalles del programa y de sus vulnerabilidades, incluida la posibilidad de sufrir ataques informáticos. Ante este riesgo, se deben adoptar las medidas de salvaguardia necesarias, incluyendo la denegación del acceso a este tipo de información* [subrayado nuestro].

En efecto, el escueto razonamiento dado por la Administración en este asunto hubiera merecido, quizás, alguna consideración de mayor entidad por parte del CTBG. Por ejemplo, no parece lógico que en las especificaciones técnicas del software implementado –que se ha de presuponer robusto– la propia Administración identifique la existencia de eventuales vulnerabilidades y, en consecuencia, los vectores de ataque posible. Entre otras cosas, porque es de suponer que tales vulnerabilidades, de existir y ser conocidas, habrían exigido la aplicación de parches o corrección de los errores por parte de la Administración. Lo contrario sería reconocer que, existiendo vulnerabilidades conocidas en el software del bono social, no se resolvieron debidamente por el Ministerio de Energía, Turismo y Agenda Digital, responsable del desarrollo de la aplicación informática.[143]

La realidad es que en ningún momento el CTBG entra a valorar si, efectivamente, el acceso al código fuente y/o a las especificaciones técnicas de las aplicaciones en cuestión posibilitan o no la identificación de vulnerabilidades concretas que puedan ser explotadas por un atacante para la propagación de código malicioso en el servicio web de la Administración concernida. Y, desde luego, no consta que, durante la tramitación de la reclamación y al amparo del art. 8.2.r) del Reglamento del Consejo,[144] se hubiera solicitado informe al Centro Criptológico Nacional (CCN-CERT), como organismo competente de la seguridad de los sistemas de las tecnologías de la información de las Administraciones públicas que procesan, almacenan o transmiten información en formato electrónico, que normativamente requieren protección, y que incluyen medios de cifra[145] o, en su caso, al órgano competente dentro del Ministerio en materia tecnologías y seguridad de la información.

Tal informe, en cambio, sí que es aportado por el Ministerio de la Transición Ecológica en el procedimiento-contencioso administrativo donde se dirime la impugnación de CIVIO a la resolución del CTBG sobre el acceso al código fuente. El informe pericial evacuado por el CCN identifica las dimensiones de la seguridad que podrían verse afectadas (integridad, disponibilidad y confidencialidad) por un acceso indiscriminado al código fuente:[146]

> *En las aplicaciones informáticas que manejan información clasificada o en aquellas mediante las cuales se ejercen potestades de las administraciones públicas, la pérdida de integridad, disponibilidad o confidencialidad de la información puede ocasionar un impacto que acepte gravemente a la seguridad nacional, a la seguridad pública o a la seguridad jurídica de los administrados.*
> *[…] El conocimiento del código fuente de una aplicación es un factor determinante que aumenta la peligrosidad de una vulnerabilidad al facilitar la elaboración del programa o "exploit" que permite su explotación.*
> *En definitiva, podemos concluir que la revelación del código fuente aumenta de una manera objetiva la severidad de las vulneraciones de cualquier aplicación informática. Si esta además maneja información clasificada o sensible de la administración con el conocimiento del código fuente aumenta el riesgo de que la explotación de las vulnerabilidades pueda afectar a la seguridad nacional, a la seguridad pública con la seguridad jurídica de los administrados.*

En cualquier caso, conviene señalar que es habitual que los Equipos de Respuesta a Incidentes de Seguridad Informática (CSIRT)[147], tanto el CCN-CERT como el INCIBE-CERT, emitan alertas de vulnerabilidades de criticidad alta sobre repositorios y librerías de código abierto que, en ocasiones, han llegado a ser aprovechadas por los hackers para lanzar sus ataques.[148] Piénsese, además, que desde el punto de vista de la seguridad, casi todos los servicios web de las Administraciones están conectados a bases de datos y presentan los mismos problemas que cualquier otra aplicación accesible por Internet, por lo que son una puerta de entrada para diferentes vulnerabilidades Web y ataques.[149]

6.2. La protección de la propiedad intelectual

Otro de los límites habituales al acceso al código fuente, o en su caso, al algoritmo subyacente, es el relativo a la protección de los derechos de propiedad intelectual al considerarse que esta clase de información son creaciones intelectuales. La invocación del límite –unas veces, directamente, otras en conexión con otros límites como la protección del *secreto comercial* o de los *intereses comerciales*– es habitual, dado que buena parte de las soluciones tecnológicas y aplicaciones implementadas por Administraciones y sector público son objeto de licitación y desarrollo por contratistas privados, argumentándose, además, que el acceso al código fuente o al algoritmo afectaría perjudicialmente a la posición competitiva de las empresas desarrolladoras en el mercado.[150]

En Francia, uno de los requisitos específicos, aunque no el único, para que el código fuente sea accesible a través del derecho de acceso es que la Administración detente los derechos de propiedad intelectual de dicho Código,[151] bien como titular originario en su condición de autora de una obra original (e.g. mediante el desarrollo *in-house*), bien como cesionaria de los derechos de explotación de la obra al amparo, respectivamente, de los arts. 111-1 y L131-3-1 del Código de Propiedad Intelectual Francés.[152] La CADA parece inclinarse por la aplicación del límite de la propiedad intelectual sólo en aquellos casos en que el código fuente ha sido desarrollado por un tercero y la entidad pública no es cesionaria en exclusiva de los

derechos de explotación. En este sentido, la Autoridad francesa ha lamentado en el *Asunto del Observatorio Francés de las Coyunturas Económicas (OFCE)* que los derechos de propiedad intelectual de un tercero puedan ser un *obstáculo* para el ejercicio del derecho de acceso a un código fuente que ha sido desarrollado con *fondos públicos* en el marco de las *misiones de servicio público* desarrolladas por la entidad requerida y hace un llamamiento a la revisión de los acuerdos contractuales al respecto.[153]

Precisamente, en su Resolución de 20 de octubre de 2020 sobre Aspectos Éticos de la Inteligencia Artificial, el Parlamento Europeo recomienda a la Comisión que los derechos de propiedad intelectual de terceros no sean un impedimento al acceso público al código, a los datos generados y a los modelos de IA desarrollados con fondos públicos en el marco de procedimientos de contratación.[154]

En el caso italiano donde la propiedad intelectual no figura expresamente ni entre las causas de exclusión (art. 24.1) ni entre los límites al derecho de acceso (art. 24.6) de la Ley n. 241/1990, el Tribunal Administrativo de la Región Lazio-Roma ha estimado que, cuando se trata de un software personalizado, desarrollado por un tercero, por encargo de la Administración y de acuerdo con las especificaciones técnicas concretas establecidas por esta última, habrá que estar a lo pactado entre las partes en el correspondiente contrato de desarrollo del software con relación a los derechos exclusivos de explotación y de uso. En consecuencia, no resulta invocable el límite del derecho de propiedad intelectual si en el contrato entre empresa desarrolladora del software y la Administración existen acuerdos específicos que atribuyan en exclusiva a la Administración la explotación económica y el uso del programa informático sin que se haya previsto reserva alguna en favor de la empresa desarrolladora. Aunque nada dice la Sentencia del T.A.R., la inoponibilidad de la propiedad intelectual como límite al derecho de acceso también debería sostenerse cuando se ejerza tal derecho con relación a desarrollos de software in-house.

Dado que las disciplinas del derecho de acceso y de la propiedad intelectual tienen finalidades diferentes (la primera, el escrutinio de la actividad pública, la segunda, la protección de los derechos morales y económicos del autor), en el contexto italiano que comentamos, el reconocimiento del derecho de acceso al software y al algoritmo subyacente implican no sólo el visionado del código fuente sino también la obtención de una copia. En todo caso, el derecho de acceso y, en su caso, a la obtención de una copia no exime al solicitante de la obligación de respetar los derechos de propiedad intelectual existentes en el uso que pretenda hacer del código fuente así obtenido.[155] Aquí, de nuevo, la Sentencia del T.A.R. Lazio-Roma delimita el alcance del derecho al acceso al software a la simple visualización y extracción de una copia excluyendo al máximo un derecho de reproducción con fines de explotación económica:

> *Debido al carácter económico exclusivo de la obra, la exhibición debe permitirse en las formas solicitadas por el interesado, es decir, visualización y extracción de una copia, entendiéndose que debe hacerse un uso adecuado de la información obtenida, es decir, un uso exclusivamente funcional de acuerdo con el interés manifestado en la solicitud de acceso que, por alegación expresa del solicitante [el sindicato], está representado por la protección de los derechos de sus afiliados, ya que éste [interés] constituye no solo la función para la cual se permite el acceso,*

> *sino al mismo tiempo también el límite del uso de los datos obtenidos, con la consiguiente responsabilidad directa de la persona con derecho de acceso hacia el titular del software*.

A diferencia del régimen jurídico italiano, en la legislación estatal y autonómica de transparencia, la protección de los derechos de propiedad intelectual e industrial de la propia Administración o de terceros sí que constituye un límite legal al derecho de acceso. Sin embargo, esto no ha impedido que la Autoridad catalana haya desestimado la aplicación de este límite en supuestos en los que la Administración reclamada es titular de los derechos de propiedad intelectual del código fuente o del algoritmo subyacente y el derecho de acceso tiene por finalidad verificar el correcto diseño del programa o del algoritmo para garantizar el cumplimiento estricto de los requisitos previstos en la norma jurídica que implementaba el sistema de decisión automatizado.[156] Pero a efectos de conciliar el acceso al código o al algoritmo subyacente y la protección de los derechos de propiedad intelectual de la Administración, la GAIP llega a la misma solución procedimental que el Derecho italiano a través del llamado *acceso condicionado*:

> *Por todo lo expuesto, se tiene que declarar el derecho de la persona reclamante a que le sea facilitado el código fuente solicitado, por correo electrónico, tal como pidió. Atendida la finalidad de control de la petición de acceso, se restringe el acceso a esta finalidad y no se permite la difusión o la utilización del código fuente para otras finalidades sin la autorización expresa de la Administración de la Generalitat"* [subrayado nuestro].[157]

De hecho, el *acceso condicionado* viene siendo una solución habitual en la casuística de la GAIP en el marco de solicitudes de acceso a la información contenida en expedientes de contratación relativa a las ofertas presentadas por los licitadores y que pueda estar protegida por derechos de propiedad intelectual e industrial o por declaraciones de confidencialidad al amparo de lo dispuesto en la legislación de contratos.[158]

Sin embargo, no siempre el acceso condicionado al código fuente y a las bases de datos asociadas resulta una solución adecuada, especialmente cuando existen evidencias de que la finalidad del acceso nada tiene que ver con el escrutinio o control del software empleado por la institución pública en el ámbito de sus funciones, sino más bien con fines de explotación del solicitante o de otros terceros. Precisamente para estos supuestos, resulta pertinente traer a colación el asunto del *software Cysgliad* de la Universidad de Bangor, donde el *Information Commissioner's Office* plantea una interesante ponderación entre el interés público en el acceso y la protección adecuada del interés comercial en la legítima explotación de los derechos de propiedad intelectual de su titular.[159]

A la hora de ponderar los derechos en conflicto, el ICO considera que, sin perjuicio del interés público en la transparencia y en la rendición de cuentas con relación al funcionamiento de un software desarrollado por una institución pública, y en particular, en el escrutinio de la calidad y debilidades de dicho software,[160] deben tenerse en consideración otros factores concurrentes: (i) el esfuerzo, tiempo e inversión en recursos humanos y materiales realizados por la entidad reclamada para el desarrollo y actualización del código fuente del software;[161] (ii) el hecho de que versiones anteriores del código fuente se hayan distribuido mediante licencias

open source o de código abierto;[162] (iii) el alto valor comercial del código fuente, por ser fuente exclusiva de financiación de la actividad realizada por la entidad reclamada en el interés público, a través de acuerdos de licencia con terceros;[163] (iv) la pérdida de confianza y credibilidad frente a aquellos terceros con los que mantiene acuerdos de licencia para la explotación comercial de sus derechos legítimos;[164] (v) la existencia de un riesgo real de que dicho código fuente pueda ser explotado por el propio solicitante o por terceros para el desarrollo de software propietario en directa competencia con el software de la entidad reclamada.[165]

En la ponderación de los intereses concurrentes en el caso concreto, la Autoridad británica finalmente considera que la divulgación de los códigos fuente objeto de la solicitud, permitiría a otros copiar un producto en el que la Universidad ha invertido sus esfuerzos, tiempo y dinero, resultando extremadamente complejo para la entidad reclamada controlar la explotación indebida que terceros pudieran hacer de sus derechos de propiedad intelectual, lo que, a su vez, pondría en peligro sus ingresos futuros y el propio sistema de autofinanciación de la Unidad desarrolladora, y conduciría a una pérdida de la confianza y del crédito obtenido en sus acuerdos con terceros.[166]

El CTBG también ha tenido ocasión de analizar la aplicación del límite relativo a la propiedad intelectual con relación a una aplicación telemática del bono social.[167] En este caso, el Consejo estatal estima procedente la aplicación del límite relativo a la propiedad intelectual al código fuente, ya que el software *puede ser protegido por el derecho de autor como obra literaria* al amparo de la normativa internacional y comunitaria sobre la protección jurídica de los programas de ordenador. En cambio, el Consejo entiende que la protección de los programas de ordenador prevista en la normativa de propiedad intelectual no se extiende ni a las especificaciones técnicas de la herramienta informática en cuestión ni al resultado de las pruebas realizadas para comprobar que dicha aplicación cumple la especificación funcional. En este sentido se argumenta que:

> *Las primeras [las especificaciones técnicas] pueden incluir aspectos, entre otros, como si es un sistema operativo de código abierto, cómo realiza el almacenamiento de datos, cuál es su lenguaje de programación o si incluye herramientas para depuración de memoria y análisis del rendimiento del software. Existen multitud de especificaciones técnicas de programas de ordenador expuestas al público en Internet. Las segundas [las pruebas] no inciden en el hardware protegido y sirven a nuestro juicio al propósito perseguido por la LTAIBG de controlar la acción pública y el proceso de toma de decisiones.*

Sin embargo, el argumento diferenciador no resulta convincente para hacer tal distinción entre el código fuente y sus especificaciones técnicas. A efectos de invocar la aplicación del límite en cuestión, cabe recordar que el art. 96.1 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), protege no sólo los programas de ordenador, *cualquiera que fuere su forma de expresión y fijación*, sino también *su documentación preparatoria*, así como *la documentación técnica y los manuales de uso de un programa*.[168]

En su recurso contencioso-administrativo, considera CIVIO que el límite del artículo 14.1.j) de la LTAIBG relativo a la protección de la propiedad intelectual o industrial sólo puede aducirse por la Administración

> *cuando el código objeto de petición pudiera ser de titularidad de un tercero, pero nunca y en ningún caso cuando la titularidad de dicho código es de una administración pública, puesto que nada hay en la legislación que permita ocultar la motivación de los actos con los que se nos gobierna. Interpretarlo de manera contraria e impedir el acceso al código fuente permitirá que la administración pública, tanto en el presente como en el futuro, desarrolle algoritmos ocultos al escrutinio público puesto, al hallarse tales algoritmos regulados por propiedad intelectual, gozarán de la excepción del artículo 14.1.j).*

Por el contrario, la Fundación recurrente considera que el código fuente desarrollado por una Administración no es objeto de propiedad intelectual según una interpretación amplia y sistemática de los arts. 13 y 31 del TRLPI.

A mayor abundamiento, se ha defendido por algunos autores que todo software utilizado por las Administraciones públicas o por los poderes del Estado para ejercer las competencias que le son propias son *de facto* una norma jurídica. En consecuencia, todas las garantías formales de las normas deberán aplicarse al código fuente. Tales garantías se concretarían en que debe existir una regulación jurídica específica que establezca tanto el procedimiento de la elaboración como los órganos competentes para la escritura del código fuente y de los algoritmos –a los que los ciudadanos también deben tener acceso, porque de ellos depende la aplicación de las normas jurídicas; los repositorios del códigos fuente y los algoritmos deben ser públicos para dar la posibilidad de que la ciudadanía pueda leerlos y realice las alegaciones correspondientes; y, finalmente, si el código fuente y los algoritmos son normas jurídicas, no pueden estar sujetos entonces a derechos de propiedad intelectual.[169]

7. Contenido y alcance del principio de transparencia algorítmica. La necesaria convergencia entre iuspublicismo y "XAI"

Las demandas de una mayor *transparencia algorítmica* de las decisiones basadas o adoptadas mediante sistemas de IA han ido creciendo en el debate público y político en la medida en que las organizaciones públicas y privadas han ido intensificando el uso de grandes volúmenes de información personal y de complejos sistemas de analítica de datos para la toma de decisiones. En este sentido, el *principio de transparencia algorítmica* con relación a los usos concretos y decisiones basadas o adoptadas por las Administraciones y sector público mediante sistemas de IA aparece bajo diversas formulaciones en las declaraciones generales y recomendaciones de organismos internacionales[170] y sociedad civil,[171] estrategias nacionales sobre IA[172] y en la literatura jurídica,[173] si bien es cierto que también es habitual el uso de otras formulaciones del principio (e.g. *transparencia de los algoritmos*, *transparencia algorítmica*, *transparencia y explicabilidad*, *transparencia y trazabilidad*, *transparencia técnica*).

Ahora bien, conviene adelantar ya que el significado *técnico* de la *transparencia algorítmica* en el ámbito de las Ciencias de la Computación no es exactamente coincidente con el significado que tiene el principio de transparencia de la acción pública en el ámbito jurídico, aunque la finalidad última en el plano técnico y el jurídico con relación a los algoritmos sea *abrir*

al conocimiento de terceros (el público y sociedad civil, en general, las autoridades de control, los expertos) las razones o criterios de una decisión basada o adoptada mediante sistemas de IA.

Es por ello que resulta imprescindible analizar el significado de la *transparencia algorítmica* en ambos planos, el estrictamente técnico y el jurídico. Es más, cualquier propuesta regulatoria que pueda hacerse en el ámbito de la aplicación de la IA, ya sea desde el sector público o privado, debe pasar necesariamente por tener en cuenta el dominio técnico, so pena de incurrir en lo que Gómez Jiménez ha denominado el *efecto de la Reina Roja* en el ámbito del Derecho.[174]

En el caso particular de los usos que Administraciones y sector público hacen de los sistemas de IA con impactos individuales o colectivos en los interesados en particular o en los ciudadanos en general este dominio técnico debe estar presente también en cualquier enfoque que se adopte desde la perspectiva de la transparencia, en su doble vertiente de publicidad activa y derecho de acceso, de la actividad pública algorítmica.

7.1. Las limitaciones de la legislación de transparencia como mecanismo de escrutinio de la algoritmia decisional de las Administraciones

Partiendo del estrecho nexo entre transparencia y rendición de cuentas, desde algunos planteamientos iuspublicistas se viene identificando la expresión *transparencia algorítmica* no sólo con ciertas obligaciones de publicidad activa, sino también con el derecho de acceso al código fuente y al algoritmo subyacente.[175] En este sentido, algunas de las propuestas comunes para regular jurídicamente los algoritmos de IA en el ámbito público vienen poniendo el acento en el *principio de transparencia algorítmica*, entendido como una exigencia a las organizaciones públicas de *exponer el código fuente y de los datos subyacentes a un cierto grado de escrutinio público.*[176]

Sin embargo, esta aproximación a la transparencia algorítmica no está exenta de críticas. Aunque, *prima facie*, parece que la transparencia total *por defecto* es la opción que mejor garantizaría la máxima rendición de cuentas de las organizaciones –públicas o privadas–, existen argumentos que cuestionan la oportunidad de este planteamiento: (1) el impacto en la privacidad, dado que datos personales confidenciales podrían filtrarse fácilmente a la luz pública; (2) efectos perversos de *gaming* o engaño de los modelos orientados al cumplimiento normativo ya que, al ser accesibles las variables *proxy*, ello permitiría eludir la aplicación de las normas (e.g. potenciales infractores podrían identificar los *red flags* determinantes de perfiles de riesgo de evasión fiscal); (3) el riesgo de prácticas contrarias a la competencia, al ser accesibles por terceros los algoritmos y el *know how* asociado a su desarrollo e implementación por decaer la protección jurídica amparada en los derechos de propiedad intelectual y los secretos comerciales frente a la apertura total del modelo; (4) la naturaleza esencialmente dinámica y adaptativa de estos modelos, que se actualizan y reajustan con la entrada de nuevos datos a lo largo del tiempo con lo que la dimensión temporal de la transparencia puede resultar problemática (5) la opacidad inherente a los modelos de decisión algorítmica, en particular, los de caja negra, caracterizados por su alto rendimiento pero baja interpretabilidad.[177]

Precisamente, este último argumento de la *opacidad inherente* es el que está teniendo mayor calado entre la doctrina jurídica: la revelación del código fuente del algoritmo de aprendizaje automatizado se considera una aproximación insatisfactoria pues el código suele

ser incomprensible para los no expertos;[178] y su análisis por un experto aportaría muy poco ya que el código sólo revelaría el método de *machine learning* utilizado, pero no la *regla decisional* existente entre los datos de entrada y los resultados del modelo[179] o por qué sistema particular tomó una decisión concreta en una situación dada.[180] De hecho, existe un consenso más o menos generalizado en que la transparencia absoluta de un modelo de aprendizaje automatizado –entendida como el acceso al código fuente– no siempre permite entender *cómo funciona un sistema complejo, qué partes son esenciales en dicho funcionamiento o hasta qué punto la efímera naturaleza de las representaciones computacionales es compatible con la legislación de transparencia.*[181]

Así, por ejemplo, al analizar en qué medida la *Freedom of Information Act* puede garantizar la transparencia de las decisiones algorítmicas, Fink considera que, si bien poner a disposición de los ciudadanos el código fuente y/o el algoritmo subyacente utilizado por una Administración puede encajar plenamente en los principios de gobierno abierto, se trata tan sólo de un paso necesario, pero no suficiente, por la sencilla razón de que la transparencia administrativa en sí misma no es sinónimo de una mejor rendición de cuentas, sino tan sólo una condición previa.[182]

Con un planteamiento similar, se han pronunciado tanto el Consejo de Europa como la Comisión Europea, para quienes la *transparencia algorítmica* no equivale exactamente a publicar o acceder al código del algoritmo.[183]

A resultas de lo anterior, la doctrina comparada viene considerando que el derecho de acceso resulta un instrumento insuficiente para garantizar debidamente la transparencia algorítmica. La legislación de transparencia –se insiste– privilegia el *ver* sobre el *entender*, pero *ver* dentro del modelo algorítmico no significa mejor comprensión de su comportamiento o de las variables determinantes de sus resultados.[184]

Ya en clave interna, nuestra doctrina también ha considerado insuficiente el derecho de acceso,[185] calificando la LTAIBG como un instrumento excesivamente restrictivo para garantizar la efectiva transparencia de los sistemas automatizados.[186] Por un lado, se subraya que nuestra legislación de transparencia actual no contempla ninguna obligación de publicidad activa similar a la prevista en el art. 45.4 de la antigua LRJAP-PAC[187], que preveía la publicación de las *características* de las aplicaciones utilizadas por las Administraciones en el ejercicio de sus potestades. Una obligación similar –se dice– permitiría a los ciudadanos conocer el funcionamiento de dichas aplicaciones e incluso impugnarlas cuando resultasen lesivas para sus derechos. Es más, la legislación de transparencia española no señala nada específico sobre la IA en el ámbito administrativo.[188] Por otro, desde la perspectiva del derecho de acceso a la información pública, y suponiendo que la ponderación de intereses resultara en la prevalencia del interés público en el acceso frente a límites como la seguridad pública, los intereses comerciales, la propiedad intelectual e industrial o incluso las eventuales declaraciones de confidencialidad del contratista privado que ha desarrollado el modelo para la Administración, el acceso al código fuente y a sus especificaciones técnicas tampoco harían más transparente al algoritmo.[189]

Ahora bien, el hecho de que las decisiones adoptadas mediante modelos de aprendizaje automatizado puedan tener naturaleza de *caja negra* no significa que tales decisiones sean totalmente incomprensibles para el examen humano, sino más bien que las técnicas de aprendizaje automático no resultan tan intuitivamente interpretables como otras formas más tradicionales de análisis de datos.[190]

Es por ello, que el principio de transparencia pública, con sus dos manifestaciones concretas, publicidad activa y derecho de acceso, necesita de una reformulación a partir de la significación y connotaciones propias que la noción de transparencia tiene en el ámbito del aprendizaje automatizado. Esto último nos sitúa en los dominios de la llamada «XAI» o «Explainable Artificial Intelligence».

7.2 Interpretabilidad, explicabilidad y transparencia de las decisiones algorítmicas desde la "XAI".

Como ya se ha venido insistiendo en páginas precedentes, en términos generales existe una relación inversa entre interpretabilidad y el rendimiento de un modelo, por lo que modelos más simples son más interpretables, pero tienen una capacidad predictiva menor; y al revés.[191] Precisamente, para resolver el problema de la interpretabilidad de los modelos de IA y, en particular, de los *modelos de caja negra*, dentro del ámbito de las Ciencias de la Computación, la rama de la IA denominada *Inteligencia Artificial Explicable* (por sus siglas en inglés, *XAI)* comprende el conjunto de técnicas de *machine learning* cuya finalidad es generar modelos más explicables manteniendo niveles altos de rendimiento.[192]

La XAI parte de la importante distinción entre los conceptos de *interpretabilidad*, *explicabilidad* y *transparencia del modelo*.

No existe un concepto matemático de la *interpretabilidad*. La *interpretabilidad* es una característica *pasiva* del modelo que se refiere al grado en que un modelo concreto es comprensible o inteligible para un observador humano. Por tanto, el aprendizaje automático interpretable se refiere a métodos y modelos que hacen que el comportamiento y las predicciones de los sistemas ML sean comprensibles para los humanos. La interpretabilidad de un modelo es mayor si resulta fácil para una persona razonar y rastrear de una forma coherente por qué el modelo hizo una predicción concreta. En términos comparativos, un modelo A esa más interpretable que otro modelo B si las predicciones de A son más fáciles de entender que las realizadas por B.[193]

En cambio, la *explicabilidad* es una característica *activa* del modelo que se refiere a la capacidad de generar una explicación sobre el comportamiento del modelo a partir de los datos utilizados, de los resultados obtenidos y del proceso completo de la toma de decisión[194] en función de la audiencia o perfil de los destinatarios a los que se dirige la explicación.[195] Las explicaciones son el medio a través del cual pueden explicarse las decisiones de un modelo de ML de una forma clara, comprensible, transparente e interpretable. Por tanto, si la interpretabilidad es el objetivo final a conseguir, las explicaciones son herramientas para conseguir la interpretabilidad del modelo.[196]

A su vez dentro la XAI se diferencia entre los modelos que son *interpretables por diseño* (*modelos transparentes*) de aquellos otros que, no siendo interpretables *prima facie*, sin embargo, pueden ser explicables mediante distintas técnicas para generar explicaciones mediante la extracción de información relevante del modelo.[197]

La transparencia de un modelo de IA viene determinada por el grado de interpretabilidad intrínseca de un modelo específico. Por tanto, desde el punto de vista técnico, la transparencia es un atributo del modelo que definiría el grado de comprensibilidad que, para

un humano, tiene dicho modelo por sí mismo. En este sentido, la transparencia del modelo se mediría en tres niveles:[198]

(i) Con relación al conjunto del modelo (*simulabilidad*). El funcionamiento puede ser reproducido o replicado por un humano en un tiempo razonable a partir de los datos y parámetros del modelo mediante los cálculos necesarios para generar la predicción. Así, por ejemplo, dentro de los modelos de regresión con buena interpretabilidad, se considera que LASSO tiene un mejor nivel de interpretabilidad que el Modelo Lineal Generalizado (GLM).

(ii) Con relación a sus componentes individuales (*descomponibilidad*). Los componentes del modelo, inputs, parámetros y cálculo, admiten una explicación intuitiva. Así, por ejemplo, el nodo en un árbol de decisión puede corresponderse con una descripción en lenguaje natural (e.g. todos los pacientes con presión distólica superior o igual a 150). De la misma manera, los parámetros en modelo lineal representan la relación o coeficiente entre cada variable predictora y la predicción.

(iii) Con relación al algoritmo de entrenamiento implementado por el modelo (transparencia algorítmica). Se refiere a la capacidad del usuario del modelo de entender el proceso seguido por el modelo para producir un resultado concreto a partir de los datos. Se considera así que los bosques aleatorios, los modelos soporte vectorial (SVM), las redes neuronales profundas (multicapa, convolucionales o recurrentes) y los métodos de ensamble son modelos opacos de interpretabilidad baja o baja, por lo que necesitan ser explicados mediante técnicas complementarias. En cambio, las reglas/listas de decisión o los árboles de decisión, siempre que las listas no sean largas o los árboles muy profundos, proporcionan los mejores niveles de interpretabilidad de todas las técnicas algorítmicas de rendimiento óptimo y no opacas.

En consecuencia, un modelo de IA se considera transparente si es interpretable por sí mismo, es decir, si el funcionamiento global del modelo, de sus componentes individuales y de su algoritmo de aprendizaje resultan inteligibles o comprensibles para un humano.[199] La transparencia general de un modelo dependerá, en todo caso, de un adecuado equilibrio entre la simulabilidad, la descomponibilidad y la transparencia algorítmica. Así, por ejemplo, en los modelos lineales de alta dimensionalidad o ingeniería de variables (*feature engineering*) compleja, su transparencia algorítmica no es controvertida, sin embargo, pero pierden su simulabilidad y descomponibilidad respectivamente.[200]

Cuando el modelo no tiene transparencia intrínseca, especialmente en el caso de los modelos de caja negra, existen métodos y técnicas para garantizar su interpretabilidad. Existe toda una variedad y taxonomía de métodos y técnicas de interpretabilidad que pueden clasificarse en atención a distintos criterios.

Uno de los criterios de clasificación más relevantes de las técnicas de interpretabilidad distingue entre *métodos intrínsecos* y *métodos post hoc.* En el caso de los métodos intrínsecos, responden a la cuestión de *cómo funciona el modelo*; la interpretabilidad se logra aplicando ciertas restricciones (e.g. dispersión, monotonicidad, causalidad) para limitar la complejidad del modelo de aprendizaje automático; y se aplican a modelos que son interpretables por sí mismos (*intrinsic*). En el caso de las técnicas *post hoc*, la interpretabilidad del modelo se obtiene aplicando métodos que analizan el modelo después del entrenamiento. Las interpretaciones *post hoc* post responden a la cuestión de *qué más puede decirnos el modelo* y representan un

conjunto de técnicas para extraer información de los modelos de aprendizaje menos interpretables. Mientras que las interpretaciones *post hoc* no aclaran con exactitud cómo funciona un modelo internamente, sin embargo, pueden aportar información relevante para los expertos y usuarios finales del modelo de aprendizaje automático sin sacrificar su rendimiento predictivo.[201]

Las técnicas *post-hoc* tratan de capturar los atributos esenciales del comportamiento observable de un modelo de *caja negra*, ofreciendo modelos de explicación diferentes: (i) determinar la sensibilidad de los resultados de un modelo de caja negra a las perturbaciones en sus *inputs* de entrada; (ii) permitir la exploración interactiva de las características de comportamiento; (iii) construir modelos sustitutos más simples e interpretables para obtener una mejor comprensión de las predicciones y clasificaciones particulares, o del comportamiento del sistema en su conjunto.[202]

Existen diversidad de técnicas *post-hoc*. Las *explicaciones en lenguaje natural* pueden consistir en herramientas textuales o visuales que proporcionan una comprensión cualitativa de la relación entre las variables de una entrada (e.g., palabras en un documento) y el resultado del modelo (e.g. una clasificación o predicción). Las *técnicas basadas en la simplificación de modelos* tratan de explicar un modelo opaco mediante la construcción de modelos subrogados o sustitutos más interpretables (normalmente, mediante árboles de decisión o reglas de decisión) a partir de los datos de entrada y las predicciones generadas por el modelo opaco en cuestión (e.g. LIME). Las *técnicas de visualización* pueden demostrar visualmente la influencia relativa de variables o proporcionar una interfaz para que los usuarios exploren explicaciones textuales o visuales (e.g. PDP e ICE). Las *técnicas basadas en la relevancia de las variables* cuyo objetivo es describir el funcionamiento de un modelo opaco clasificando o midiendo la influencia, relevancia o importancia que cada variable ha tenido en los resultados del modelo (e.g. SHAP, contrafácticos). Los *técnicas basadas en ejemplos* seleccionan instancias particulares del conjunto de datos para explicar el comportamiento del modelo de aprendizaje automático o para explicar la distribución de datos subyacente.[203]

Otra clasificación importante es la que distingue entre métodos basados en *modelos específicos* y en *modelos agnósticos*. Los métodos basados en modelos específicos están limitados a técnicas de interpretación de modelos concretos que se centran en el análisis de algunas partes internas (e.g. la interpretación de los coeficientes o pesos en un modelo lineal). Los métodos basados en modelos agnósticos son aplicables a cualquier modelo ML, ya sea *black box* o no, y se llevan a cabo una vez que el modelo ha sido entrenado. Estos métodos se basan en el análisis de pares de variables de entrada y resultado y, por definición no tienen acceso a las operaciones internas del modelo, como los pesos o la información estructural.[204]

Asimismo, las interpretaciones pueden ser *locales* o *globales*. Las interpretaciones locales tienen como objetivo interpretar predicciones o clasificaciones individuales correspondientes a instancias concretas con el fin de identificar las variables de entrada específicas que han podido ser determinantes o han tenido más peso en la generación de un predicción o clasificación particular. Por su parte, las explicaciones globales tienen como objetivo ofrecer una visión amplia que abarque la importancia general de las variables y de sus interacciones en los resultados generados por el modelo, el funcionamiento interno y la lógica del comportamiento de ese modelo en su conjunto. Las interpretaciones globales se centran en explicar el conjunto

del modelo, y no tanto en su rendimiento, para un caso particular, y pueden contribuir a que el proceso de toma de decisiones sea coherente desde el punto de vista procedimental.[205]

De lo explicado anteriormente puede concluirse que, desde un plano estrictamente técnico, no hay un concepto unívoco de *transparencia algorítmica*, sino que se diferencia entre la *transparencia íntríseca* del modelo, la *transparencia algorítmica en estricto sentido* y los *métodos o técnicas de interpretabilidad* que posibilitan, en mayor o menor medida, hacer más comprensible y *más transparente* al observador humano el funcionamiento global de un modelo opaco, de sus componentes individuales y de su algoritmo de aprendizaje mediante explicaciones complementarias.

7.3. Transparencia técnico-algorítmica *vs* transparencia pública

Si se compara el significado técnico de la transparencia algorítmica con el significado jurídico del principio de transparencia en el ámbito del iuspublicismo veremos que los conceptos no son coincidentes, aunque la finalidad última, en ambos casos, sea conocer cómo se adoptan las decisiones en un determinado ámbito y, en última instancia, garantizar la rendición de cuentas. Mientras que el concepto técnico de la transparencia algorítmica podríamos decir que es polisémico y el sujeto destinatario de la misma puede ser distinto (el usuario del sistema, las autoridades de control, los expertos, el público en general), sin embargo, el significado jurídico del principio de transparencia es unívoco y el destinatario final siempre es la ciudadanía, depositaria última del derecho a saber qué hacen sus instituciones para someterlas al escrutinio público y al control democrático.[206]

En efecto, el fin último de las leyes de transparencia es hacer efectivo el derecho de los ciudadanos a *conocer cómo se toman las decisiones que les afectan, cómo se manejan los fondos públicos o bajo qué criterios actúan nuestras instituciones*. En la práctica, este *derecho a saber* posibilita y garantiza que los ciudadanos puedan *juzgar mejor y con más criterio la capacidad de sus responsables públicos y decidir en consecuencia*, contribuyendo así a una *mejor fiscalización de la actividad pública*[207] y, en suma, a la debida rendición de cuentas de las instituciones públicas, de sus representantes y gestores.[208]

Si trasladamos esta conceptualización teleológica de la legislación de transparencia al ámbito de la algoritmia decisional de la Administración, podríamos afirmar que, como un principio de orden sustantivo, la transparencia pública, en sus dos vertientes, de publicidad proactiva y del derecho de acceso, *en teoría*, debería posibilitar a los ciudadanos saber y comprender: (i) en qué ámbitos concretos de la actividad pública se adoptan decisiones basadas total o parcialmente en tratamientos algorítmicos; (ii) en qué medida esas decisiones algorítmicas afectan a los derechos y libertades de los ciudadanos, individual o colectivamente; (iii) cómo y bajo qué criterios se han adoptado esa clase de decisiones por parte de las instituciones públicas; (iv) y, por su incidencia en la racionalidad y eficiencia del gasto público, cómo se han adquirido e implementado esos sistemas de decisión algorítmica (e.g. mediante desarrollos *in house* o mediante licitación), qué fines institucionales específicos se pretenden cumplir, qué necesidades públicas satisfacer y por qué dicha implementación tecnológica innovadora es la mejor alternativa frente a otras soluciones.

De hecho, algunas de las propuestas regulatorias comienzan a plantear una noción del principio transparencia muy próxima a la noción de interpretabilidad manejada en el ámbito técnico. Así, por ejemplo, la Resolución del Parlamento Europeo de 16 de febrero de 2017, con

recomendaciones destinadas a la Comisión de normas de Derecho civil sobre robótica define *el principio de transparencia en el sentido de que*

> *siempre ha de ser posible justificar cualquier decisión que se haya adoptado con ayuda de la inteligencia artificial y que pueda tener un impacto significativo sobre la vida de una o varias personas; […] siempre debe ser posible reducir los cálculos del sistema de inteligencia artificial a una forma comprensible para los humanos [subrayado nuestro].*[209]

Desde esta perspectiva, la transparencia posibilitaría la necesaria rendición de cuentas de las instituciones públicas que implementan sistemas algorítmicos en sus procesos de toma de decisiones.[210] A su vez, la rendición de cuentas debería garantizar que las instituciones expliquen y justifiquen las decisiones adoptadas mediante sistemas algorítmicos[211] de una manera transparente, motivada y comprensible para sus destinatarios; generen evidencias documentales suficientes, confiables, íntegras y trazables que permitan supervisar y verificar el correcto funcionamiento del modelo algorítmico, con la finalidad de poder imputar y exigir la responsabilidad correspondiente por las decisiones adoptadas;[212] y, desarrollen y usen de forma responsable los sistemas algorítmicos, garantizando el respeto a los derechos humanos y contribuyendo a generar beneficio social.[213]

Sin embargo, no puede ignorarse que, cuando la LTAIBG y legislación autonómica concordante establecen que la información sujeta a obligaciones de publicidad activa se publique de forma *entendible* y *comprensible* para los ciudadanos, en realidad, se trata tan sólo de una exigencia instrumental y al servicio de los principios técnicos de organización de los Portales de Transparencia a efectos de garantizar la *accesibilidad universal* y la *usabilidad* de estos portales.[214] A esto se une que, desde la perspectiva del derecho de acceso, quedan excluidas del concepto de información pública las solicitudes que compelen a la Administración a motivar adicionalmente las resoluciones más allá de la fundamentación que se haya utilizado o las aclaraciones o pronunciamientos concretos sobre una determinada resolución,[215] por lo que parece claro que ni la publicidad activa ni el derecho de acceso son el instrumento adecuado para garantizar la total comprensibilidad o la inteligibilidad de la actividad administrativa algorítmica, especialmente, si en el proceso de toma de decisiones puede estar involucrado un algoritmo de IA. En todo caso, tanto la publicidad activa como el derecho de acceso podrán coadyuvar a constatar la concurrencia o ausencia de razones de una determinada actuación.

En definitiva, la exigencia de *comprensibilidad* invocada por la legislación de transparencia tan sólo es eso, una condición instrumental para la consecución de los fines propios de la norma, pero que, en la práctica, no garantiza por sí misma que el ciudadano pueda entender plenamente *bajo qué criterios actúan las instituciones públicas* cuando adoptan sus decisiones (total o parcialmente) mediante procedimientos algorítmicos. Más concretamente, el derecho de acceso no garantiza en sí mismo la racionalidad y justificación intrínseca de las decisiones públicas algorítmicas (pero, en realidad, tampoco las de las decisiones plenamente humanas), porque eso ya corresponde al dominio de la *motivación* de la decisión administrativa.

Pero llegado a este punto nos encontramos con un nuevo problema. Teniendo en cuenta los conceptos técnicos de transparencia, interpretabilidad y explicabilidad previamente analizados, es posible afirmar que el deber de motivación de los actos administrativos recogido

en el art. 35.1 en la Ley 39/2015 resulta a todas luces insuficiente, al menos, en los términos de su vigente formulación (*sucinta referencia de hechos y fundamentos de derecho*) para garantizar la transparencia (en su polisémico significado de simulabilidad, descomponibilidad y transparencia algorítmica en estricto sentido), y en su caso, la explicabilidad, de las decisiones basadas o adoptadas en sistemas de aprendizaje automatizado. De hecho, no existe en nuestra legislación administrativa una obligación explícita que compela a la Administración a *explicar cómo se ha llegado a una decisión (algorítmica), cuál ha sido el proceso, qué datos se han tenido en cuenta, qué objetivos se persiguen*.[216]

7.4. Necesidad de redefinir el contenido y alcance de la transparencia algorítmica desde el iuspublicismo

Al analizar cómo se está trasladando el principio de transparencia pública al ámbito de las decisiones basadas o adoptadas mediante el uso de sistemas de IA por parte de Administraciones e instituciones, la mayoría de las propuestas encontradas en el ámbito internacional y comparado, en su mayoría procedentes del ámbito de la *soft law*, se vienen limitando a enunciados generales del principio.

Si bien es cierto que el principio de transparencia suele anudarse a los principios de *explicabilidad*, *trazabilidad* y/o *auditabilidad*, estas propuestas o recomendaciones se caracterizan por: (i) un alcance subjetivo muy limitado respecto de la ciudadanía en general;[217](ii) por la inconcreción del contenido y alcance objetivo del principio (e.g. qué obligaciones de información concreta implicaría); (iii) la ausencia de restricciones al uso de modelos de *black box* o, al menos, la imposición de requerimientos específicos al uso de estos modelos que garanticen su transparencia (e.g. explicaciones *ex post* que permitan comprender las posibles consecuencias adversas del uso de estos modelos).

Sería el caso de la Recomendación de la OCDE sobre Instrumentos Legales para la Inteligencia Artificial de 2021,[218] la Recomendación del Consejo de Europa CM/Rec(2020)1, del Comité de Ministros a los Estados Miembros sobre el impacto en los derechos humanos de los sistemas algorítmicos;[219] la Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas de 20 de octubre de 2020;[220]; las Directrices Éticas para una IA Fiable del Grupo de Expertos de Alto Nivel para la IA de la Comisión Europea de 2018;[221] o la Carta de Derechos Digitales de la Secretaría de Estado de Digitalización e Inteligencia Artificial cuyo texto final está pendiente de aprobación tras la conclusión del proceso de consulta pública en enero de 2021.[222]

Asimismo, al analizar buena parte de los trabajos previos de la UE a la propuesta de Reglamento sobre IA, se observa la invocación constante del principio de transparencia, junto con los de trazabilidad y explicabilidad, como garantía efectiva de los derechos fundamentales de los ciudadanos, especialmente, cuando se trata de sistemas de IA considerados de *alto-riesgo*. Una vez más, estas proclamaciones no terminan de concretar las obligaciones jurídicas que tales principios implicarían para los usuarios de los sistemas de IA –ya sean instituciones públicas o privadas– respecto de los destinatarios (personas individuales o ciudadanía en general).

Así, por ejemplo, en el Libro Blanco sobre IA de la Comisión Europea, se constata, por un lado, que *la falta de transparencia (opacidad de la IA) hace difícil detectar y demostrar los posibles incumplimientos de la legislación, especialmente las disposiciones legales que protegen*

los derechos fundamentales, imputan responsabilidades y permiten reclamar una indemnización; y, por otro, que en los regímenes jurídicos de la UE no se contemplan de manera específica requisitos de transparencia. De lo anterior concluye la Comisión que *la opacidad de los sistemas basados en algoritmos puede abordarse mediante requisitos de transparencia*. Sin embargo, al abordar las obligaciones de transparencia exigidas a los usuarios de los sistemas de IA respecto de los potenciales destinatarios o afectados por dichos sistemas, se advierte que las propuestas de la Comisión pasan prácticamente por alto esta cuestión, lo cual resulta llamativo en el caso de que los usuarios de los sistemas de IA sean las Administraciones y autoridades públicas en la medida en que se encuentran sujetos al principio de transparencia tanto en el Derecho de la Unión como en el Derecho nacional de los Estados miembros.

Se trata, por tanto, de un enfoque incoherente y limitado de la transparencia, a pesar de la constante apelación al principio. En primer lugar, porque las obligaciones de conservación de registros y datos y de suministro de información parecen reducirse exclusivamente a la rendición de cuentas frente a las autoridades de control competentes que, en su caso, se creen. En segundo lugar, porque el suministro de información relevante (en particular, las relativas a capacidades y limitaciones del sistema de IA, su objetivo o finalidad, las condiciones de funcionamiento y el nivel de exactitud esperado) se contemplan principalmente en el ámbito de las relaciones horizontales entre el diseñador del sistema y el usuario del mismo, y de forma muy secundaria, con relación a las *partes afectadas*. Por último, la única obligación de transparencia general que se plantea de forma explícita se refiere al deber de informar *claramente a los ciudadanos de cuándo están interactuando con un sistema de IA y no con un ser humano*. Respecto de esta última obligación, aunque la Comisión considera que serían necesarios *requisitos adicionales*, de nuevo, deja sin concretar cuáles serían tales requisitos, más allá de señalar que la información facilitada deberá ser *objetiva, concisa y fácilmente comprensible* y que habrá de adaptarse al contexto específico, evitándose, desde luego, *cargas innecesarias* a los destinatarios de las obligaciones de información.[223]

Este planteamiento limitado de la transparencia se ha trasladado, sin duda, a la propuesta de Reglamento Europeo para la IA. En efecto, las obligaciones de transparencia establecidas por la propuesta de Reglamento se circunscriben exclusivamente a los *sistemas de alto-riesgo*,[224] por un lado; y a los sistemas de interacción persona-máquina, sistemas de reconocimiento emocional y categorización (social) biométrica y a los sistemas de generación o manipulación de contenidos (*deep fakes*), por otro. En el caso de las obligaciones de transparencia previstas para los sistemas de alto-riesgo, si bien es cierto, que pueden parecer exhaustivas,[225] al analizar con detenimiento la propuesta se advierte en seguida que tienen un alcance subjetivo muy limitado. Así, por ejemplo, con relación a las obligaciones de suministro de información previstas en la propuesta de Reglamento,[226] tales obligaciones parecen imponerse exclusivamente a los diseñadores y desarrolladores de sistemas de IA y sus destinatarios finales serían, en todo caso, los usuarios del sistema (las organizaciones públicas o privadas que los implementen), pero no los afectados (individuales o colectivos) por las decisiones algorítmicas adoptadas con base en o mediante estos sistemas, y de ninguna manera, el público en general (cfr. art. 13.1).[227]

Asimismo, se impone un deber genérico de información a las personas individuales con relación a los sistemas de interacción persona-máquina, sistemas de reconocimiento emocional y categorización biométrica, así como sistemas de generación o manipulación de imagen, audio, o contenido de video que puedan dar lugar a *deep fake news.* Sin embargo, no se determina ni

el contenido ni el alcance de las obligaciones de información y, en todo caso, este deber de información decaería cuando esta clase de sistemas estuviesen autorizados por Ley para fines de detección, prevención, investigación y persecución de delitos. En el caso de los sistemas de generación y manipulación de contenidos, además del supuesto anterior, tampoco existirá una obligación de información al público en general cuando tales sistemas sean "*necesarios para el ejercicio de la libertad de expresión y el derecho a la libertad de las artes y de las ciencias garantizado en la Carta de Derechos Fundamentales de la Unión Europea […] con sujeción a garantías apropiadas para los derechos y libertades de tereceros*". Garantías que, en ningún momento, la propuesta de Reglamento concreta.

Además de la propuesta de Reglamento, nivel nacional europeo, existen ya algunas propuestas generales –aunque en ciertos casos, sin suficiente definición aún– que conectan la transparencia de las decisiones basadas o adoptadas mediante sistemas de IA con las obligaciones derivadas de la legislación de transparencia y acceso a la información pública.

Así, por ejemplo, en su posición común sobre la *Transparencia de la Administración Pública en el Uso de Algoritmos como Elemento Esencial para la Protección de los Derechos Fundamentales* (2018), la Autoridad Federal de Protección de Datos y de Libertad de Información y otras ocho Autoridades homónimas de los Länder han subrayado la importancia de que las entidades públicas garanticen la suficiente transparencia sobre los algoritmos utilizados, poniendo a disposición de los ciudadanos información relevante y comprensible sobre esta clase de tratamientos *tan ampliamente como sea legalmente posible*. Esta información relevante comprendería: (i) las categorías de los datos de entrada y los resultados de los datos de salida del modelo; (ii) la lógica contenida en éste y, en particular, las fórmulas de cálculo utilizadas, incluyendo la ponderación aplicada a los datos de entrada, el conocimiento especializado subyacente y la configuración individual utilizada por los usuarios; (iii) el alcance de las decisiones basadas en el modelo utilizado y los posibles efectos de las decisiones. Tales obligaciones de publicidad activa se completarían con la exigencia de obligaciones de registro y documentación de los procesos, de los parámetros esenciales, de las medidas técnicas y organizativas, y de las evaluaciones y controles de calidad periódicos realizados. Asimismo, y con el fin de garantizar una verificabilidad completa, el código fuente y, si es necesario, información relevante adicional sobre los algoritmos o los procesos de IA también deberían estar disponibles para las autoridades de control y publicarse siempre que sea posible.[228]

Por su parte, la Comisión de Ética de los Datos («Datenethikkommission») ha considerado que, cuando el uso por parte del Estado de sistemas algorítmicos tenga impactos sociales lesivos o sea relevante para la formación de la opinión pública, debería garantizarse el derecho de acceso a la información de los ciudadanos e imponerse determinadas obligaciones de publicidad. Aunque la Datenethikkommission no concreta cuál habría de ser el contenido y alcance de las obligaciones de publicidad activa o del derecho de acceso a la información relevante, sin embargo, asume el planteamiento realizado por las Autoridades de Protección de Datos y Transparencia. [229]

Otro ejemplo interesante en el ámbito legislativo es la legislación francesa de procedimiento administrativo. Al tiempo que el art. L300-2 del Código de Relaciones entre el Público y la Administración califica el código fuente utilizado por una Administración como

documento administrativo, según hemos visto *supra*, el art. L311-3-1 del mismo texto legal dispone que:

> *[…] las decisiones individuales adoptadas sobre la base de un tratamiento algorítmico incluirán una mención explícita al informar a la parte interesada. Las reglas que definen este tratamiento y las características principales de su implementación serán comunicadas por la Administración a la parte interesada si así lo solicita.*

Es decir, cuando la Administración dicte una resolución basada en un tratamiento algorítmico tiene la obligación de informar expresamente al interesado de la existencia del mismo. El interesado, por tanto, tiene derecho a conocer si ha sido objeto de algún tratamiento algorítmico.

Este último precepto ha sido objeto de desarrollo reglamentario por el Consejo de Estado francés.[230] En concreto, el art. R311-3-1-1 del CRPA establece que la decisión administrativa individual deberá contener una *mención explícita* a la *finalidad perseguida por el tratamiento algorítmico*. Ello implica el *derecho a obtener la comunicación de las reglas que definen el tratamiento y sus principales características de aplicación, así como las modalidades de ejercicio de este derecho a la comunicación y de revisión, si corresponde, ante la Comisión de Acceso a los Documentos Administrativos*. A su vez, el art. R311-3-1-2 concreta la información específica que debe facilitarse al interesado que haya sido objeto de una decisión individual fundamentada en un tratamiento algorítmico en caso de que el interesado así lo reclame. Esta información deberá facilitarse de una *forma inteligible* y deberá incluir:[231]

> *1º El grado y el modo de contribución del tratamiento algorítmico a la toma de decisión; 2º Los datos tratados y sus fuentes; 3º Los parámetros de tratamiento, y si procede, su ponderación, aplicados a la situación de interesado; 4º Las operaciones efectuadas por el tratamiento.*

Del marco jurídico previsto por el legislador francés deben destacarse tres notas importantes. En primer lugar, el objeto de la regulación francesa se centra exclusivamente en el ámbito del procedimiento administrativo y en la ampliación de los derechos y garantías del interesado dentro del procedimiento, en este caso, el derecho del interesado a saber que la decisión individual adoptada por la Administración está basada en un tratamiento algorítmico. Este planteamiento no ha estado exento de alguna crítica por parte de la doctrina francesa por excluir entre las categorías de posibles interesados a los terceros titulares de derechos e intereses legítimos colectivos, como pueden ser los expertos en computación o asociaciones de consumidores, en la medida en que su competencia técnica les permite examinar el impacto del sistema algorítmico y, en su caso, ejercer acciones colectivas en los Tribunales a decisiones algorítmicas injustas y arbitrarias.[232]

En segundo lugar, y este aspecto debe valorarse muy positivamente, a la hora de reconocer este derecho del interesado, el legislador francés no ha discriminado entre algoritmos deterministas o de aprendizaje automatizado. Este matiz anterior es importante porque, entre nuestra doctrina, hay quienes consideran que respecto de los algoritmos no predictivos, que sirven para automatizar o hacer más eficiente la aplicación de las normas, no resultaría necesario *dar a conocer al ciudadano que la decisión se ha elaborado con la ayuda de una aplicación*

informática ni, con carácter general, poner a su disposición el programa ("código-fuente"); pues, a diferencia de los algoritmos predictivos, estas aplicaciones no tienen influencia sobre el contenido de la decisión administrativa, y por ello, resultaría suficiente un control *ex ante* por parte de la Administración antes de su puesta en funcionamiento. No obstante, se reconoce también que pueden existir supuestos problemáticos (e.g. concursos de traslados de funcionarios, selección de miembros de tribunales evaluadores), en los que resulte *difícil "replicar" la aplicación de la norma sin el algoritmo, para compararla con el resultado que éste arroja*. En tales supuestos, para la impugnación de la decisión administrativa y el correspondiente control de legalidad, sí que resultaría preciso conocer el algoritmo para saber si puede haber introducido algún contenido no previsto en la norma, distorsionando su aplicación o determinando una decisión contraria a Derecho.[233]

En segundo lugar, otro aspecto fundamental de la legislación francesa analizada es que este derecho de acceso a la información relativa a los tratamientos algorítmicos que sirvan de fundamento de las decisiones administrativas individuales puede ser ejercido tanto por las personas físicas como por las personas jurídicas, más allá de las limitaciones que plantea el alcance del deber del responsable de proporcionar a los interesados información significativa sobre la lógica algorítmica en el Reglamento General de Protección de Datos en el ámbito de las decisiones automatizadas, con o sin elaboración de perfiles, dado que su aplicación sólo se circunscribe a las personas físicas y, en todo caso, tal deber de información no sería aplicable a las decisiones individuales, incluida la elaboración de perfiles, parcialmente automatizadas donde haya intervención humana significativa.[234]

Desde el ámbito de la sociedad civil también se han hecho propuestas conectadas con las finalidades propias de la legislación de transparencia y derecho de acceso. Así, entre las recomendaciones dirigidas a las Administraciones y sector público con relación a los usos de la IA que implementen, la organización civil *Access Now* incluye garantizar el principio de transparencia y explicabilidad en las decisiones adoptadas mediante esta clase de sistemas. El principio implicaría la *máxima transparencia posible* para cualquier sistema de IA empleado por el sector público. Ello exige la transparencia con relación al propósito y a cómo se utiliza y cómo funciona a lo largo del ciclo de vida del sistema. Se considera, asimismo, que *los acuerdos de confidencialidad y otros contratos con terceros bajo el pretexto de proteger la propiedad intelectual son una violación de este principio porque impiden la supervisión pública y la rendición de cuentas*. En cuanto al contenido y alcance del principio de transparencia y explicabilidad, este se concretaría en: (i) la realización de informes periódicos de dónde y cómo se utilizan y gestionan los sistemas de inteligencia artificial por las Administraciones; (ii) el uso de estándares de datos abiertos tanto en los datos de entrenamiento como en el diseño del código fuente en la mayor medida posible, con adhesión a los estándares de privacidad; (iii) realización de auditorías independientes de sistemas y datos; (iv) la elaboración de informes claros y accesibles sobre funcionamiento de cualquier sistema de IA, proporcionando información significativa sobre cómo se obtienen los resultados y qué medidas se adoptan para minimizar los impactos lesivos a los derechos; (v) la comunicación (notificación) personal cuando un sistema de IA de la Administración adopta una decisión individual que afecta a los derechos de un individuo; (vi) evitar los sistemas de caja negra.[235]

En clave ya interna, una simple revisión de nuestra literatura iuspublicística pone de manifiesto la variedad de propuestas que, con mayor o menor amplitud, plantean también obligaciones de publicidad activa.

Así, por ejemplo, con relación a los tratamientos algorítmicos que elaboren perfiles individuales, Tomás de la Quadra plantea la publicación de la correspondiente

> *memoria explicativa del fin y de los objetivos que se pretenden, así como la enumeración de las variables relevantes –12, por ejemplo– que tengan un peso conjunto más significativo […] Añadir una memoria previa, general e igual para todos en que se exprese la finalidad y el objeto del algoritmo provee de un criterio funcional de crítica al mismo que puede ser relevante, sin afectar al secreto empresarial*.[236]

Para Ponce Solé, los algoritmos y códigos fuente que se utilizan para actividad automatizada o semi-automatizada (en apoyo de las decisiones administrativas) constituyen información pública cuyo conocimiento es relevante para garantizar la transparencia de su actividad relacionada con el funcionamiento y control de la actuación pública a los efectos del art. 5.1 de la LTAIBG, *por lo que deberían publicarse en los portales de transparencia, en su caso, con los límites a que hace referencia el art. 5.3*.[237]

Por su parte, Valero Torrijos defiende el reconocimiento expreso del derecho de los ciudadanos a obtener aquella información que posibilite conocer, entre otros extremos: (i) la identificación de los programas y aplicaciones utilizados; (ii) la determinación del órgano competente para el control y supervisión del funcionamiento de la aplicación o del sistema informático; (iii) el acceso al resultado de la aplicación que afecte al interesado, el origen de los datos empleados y su naturaleza, el alcance del tratamiento realizado y el funcionamiento de la aplicación, es decir, cómo a partir de los datos de origen se llega a un determinado resultado; (iv) el conocimiento de la información anterior que no solamente debe referirse a los actos resolutorios sino también a los actos de trámite y, en particular, a los informes, borradores y documentación complementaria.[238]

Conclusiones. Algunas propuestas a modo *de lege ferenda*

El desarrollo e implementación de tecnologías disruptivas, como la Inteligencia Artificial y el Big Data, por parte de nuestras Administraciones, es un hecho innegable. En particular, las Administraciones y el sector público vienen implementando desde hace tiempo algoritmos de aprendizaje automatizado en distintos ámbitos de la actividad pública. A diferencia de la programación tradicional, donde el algoritmo establece las reglas necesarias de una forma determinista para procesar los datos de entrada y obtener unos resultados concretos, en el caso de los modelos de aprendizaje automático el sistema recibe tanto los datos de entrada como los resultados esperados, a fin de extraer las reglas o lógica que rige la relación entre los inputs y los outputs, de manera que una vez obtenidas dichas reglas, el modelo las generaliza para aplicarlas a nuevos datos de entrada y producir nuevos resultados.

Tanto en el ámbito interno como en el comparado, resulta generalizada la ausencia de un mapa claro y completo que identifique los sistemas automatizados de IA que están contratando o implementando mediante desarrollos *in house* las Administraciones y el sector público en general. En el caso de la Administración pública española no es posible conocer en qué medida la contratación pública de soluciones de IA prevalece o no sobre desarrollos *in*

house. Tampoco puede constatarse si existen o no evaluaciones de impacto previas o sistemas de gestión del riesgo que analicen *ex ante* los riesgos para los derechos y libertades individuales o colectivos que tales soluciones pueden implicar.

En este contexto, cada vez son más frecuente las demandas de ciudadanos y sociedad civil reclamando la publicidad o el acceso al código fuente o a los algoritmos subyacentes implementados por las aplicaciones y sistemas de las Administraciones para la toma de decisiones. De hecho, un análisis de la casuística comparada y española sobre el derecho de acceso a la información pública permite conocer determinados "usos inquietantes" de modelos algorítmicos de IA en la toma de decisiones administrativas o de apoyo a la adopción de las mismas.

El análisis de dicha casuística permite extraer dos conclusiones importantes. Por un lado, el código fuente de las aplicaciones utilizadas por la Administración y los algoritmos subyacentes tienen consideración de "información pública" en los términos previstos por la legislación de transparencia, sin perjuicio de los eventuales límites aplicables al acceso a tal información para la protección de otros intereses públicos o privados, como la seguridad pública o los derechos de propiedad intelectual e industrial de terceros. Por otro, que existen riesgos inherentes a la automatización, total o parcial, de la actividad administrativa (formalizada o no). Estos riesgos pueden identificarse con: (i) la existencia de reglamentación oculta, *extra* y *contra legem*, errores o *bugs* embebidos en los modelos automatizados o semi-automatizados de toma de decisiones; (ii) la *vis expansiva del black box* decisional y el riesgo de arbitrariedad administrativa, pues, en la mayoría de los casos analizados, el foco de la atención de la Autoridad administrativa revisora o del juez no es tanto si el modelo algorítmico cuestionado cuyo conocimiento y examen se pretende responde a una arquitectura determinista o de aprendizaje automatizado, sino si el interesado ha tenido capacidad de conocer cómo el sistema ha llegado a una decisión concreta y por qué; (iii) la complacencia de los responsables públicos con los resultados del modelo sin que exista un proceso de validación adecuado, supervisión humana, y de control posterior; (iv) o el impacto lesivo en derechos y libertades de los destinatarios de esas decisiones y de la sociedad en general. Existiría, por tanto, un evidente interés público en el derecho de acceso al código fuente y a los algoritmos subyacentes empleados por las Administraciones en su toma de decisiones, en la medida que el acceso permitiría el escrutinio ciudadano de la algoritmia decisional, de sus riesgos y de sus impactos individuales y sociales.

Sin embargo, existe un consenso amplio en que la transparencia absoluta, bien mediante obligaciones de publicidad activa, bien mediante el derecho de acceso al código fuente del algoritmo o a toda su documentación técnica tampoco garantizaría la transparencia de las decisiones adoptadas con base en o mediante sistemas de IA, y en particular, los de caja negra, en el sentido de que no permitiría al ciudadano saber y comprender con qué criterios se han adoptado tales decisiones.

A lo anterior debe añadirse que, cuando se habla de transparencia algorítmica desde el punto de vista técnico su significado no es exactamente coincidente con el concepto jurídico del principio de transparencia en el ámbito del iuspublicismo, aunque la finalidad última de ambos conceptos (el técnico y el jurídico), sea conocer cómo se adoptan las decisiones en un determinado ámbito y, en última instancia, garantizar la rendición de cuentas.

Mientras que el concepto técnico de la transparencia algorítmica podríamos decir que es polisémico (transparencia intrínseca propia de los modelos interpretables o comprensibles, la transparencia algorítmica en sentido estricto y la transparencia resultante de la explicabilidad

de los modelos mediante técnicas complementarias de la interpretabilidad) y el sujeto destinatario de la misma puede ser distinto (el usuario del sistema, las autoridades de control, los expertos, el público en general), sin embargo, el significado jurídico del principio de transparencia es unívoco y el destinatario final siempre es la ciudadanía, depositaria última del derecho a saber qué hacen sus instituciones para someterlas al escrutinio público y al control democrático.

Si en el ámbito técnico, un modelo de IA se considera transparente cuando el funcionamiento global del modelo, de sus componentes individuales y de su algoritmo de aprendizaje resultan inteligibles o comprensibles para un humano, en el caso de la legislación de transparencia, tal comprensibilidad o inteligibilidad de los modelos no puede garantizarse mediante obligaciones de publicidad activa o mediante el derecho de acceso, a menos que se imponga *ex ante* un deber explícito a las Administraciones de tener en su poder y conservar debidamente documentación técnica específica que garantice la plena interpretabilidad y explicabilidad de los modelos de IA implementados en su actividad. En el caso de nuestro ordenamiento jurídico, la exigencia de "comprensibilidad" de la información pública invocada por la legislación de transparencia tan sólo es una condición instrumental para la consecución de los fines propios de la norma, pero en realidad tampoco garantiza por sí misma que el ciudadano pueda entender plenamente *bajo qué criterios actúan las instituciones públicas* cuando adoptan sus decisiones (total o parcialmente) mediante procedimientos algorítmicos.

Dada la ausencia de una regulación específica de la IA en el ámbito de las Administraciones y del sector público, puede afirmarse que, *de lege lata*, nuestra legislación vigente de transparencia no garantiza en sí misma un adecuado escrutinio público la racionalidad y justificación intrínseca de las decisiones públicas algorítmicas, porque eso ya corresponde al dominio de la "motivación" de la decisión administrativa.

Sin embargo, a la vista de los conceptos técnicos de transparencia, interpretabilidad y explicabilidad analizados, es posible afirmar que el deber de motivación de los actos administrativos recogido en el art. 35.1 de la Ley 39/2015 resulta insuficiente, al menos, en los términos de su vigente formulación ("sucinta referencia de hechos y fundamentos de derecho") para garantizar la transparencia técnica necesaria (simulabilidad, descomponibilidad y transparencia algorítmica en estricto sentido), y en su caso, la explicabilidad de las decisiones basadas o adoptadas en sistemas de aprendizaje automatizado. Es más, no existe en nuestra legislación administrativa una obligación explícita que compela a la Administración a explicar cómo se ha llegado a una decisión algorítmica, cuál ha sido el proceso, qué datos se han tenido en cuenta, qué objetivos se persiguen.

Al analizar cómo se está trasladando el principio de transparencia pública al ámbito de las decisiones basadas o adoptadas mediante el uso de sistemas de IA por parte de Administraciones e instituciones públicas, la mayoría de las propuestas encontradas en el ámbito internacional y comparado, en su mayoría procedentes del ámbito de la *soft law*, se vienen limitando a enunciados generales del principio, caracterizados por: (i) un alcance subjetivo muy limitado respecto de la ciudadanía en general; (ii) por la inconcreción del contenido y alcance objetivo del principio (e.g. qué obligaciones de información concreta implicaría); (iii) o, la ausencia de restricciones al uso de modelos de *black box*, o al menos la imposición de requerimientos específicos al uso de estos modelos que garanticen su transparencia (e.g. explicaciones *ex post* que permitan comprender las posibles consecuencias

adversas del uso de estos modelos). En nuestro ordenamiento, tal sería el caso de la Carta de Derechos Digitales de la Secretaría de Estado de Digitalización e Inteligencia Artificial cuyo texto final está pendiente de aprobación tras la conclusión del proceso de consulta pública en enero de 2021.

Por su parte, con relación a las obligaciones de transparencia y suministro de información previstas en la propuesta de Reglamento europeo sobre IA, tales obligaciones parecen imponerse exclusivamente a los diseñadores y desarrolladores de sistemas de IA, donde los destinatarios finales de tales deberes de transparencia serían, en todo caso, los usuarios del sistema (las organizaciones públicas o privadas que los implementen) o las Autoridades de control, pero no los afectados (individuales o colectivos) por las decisiones algorítmicas adoptadas con base en o mediante estos sistemas y de ninguna manera el público en general.

Cualquier regulación que se realice del uso de la IA en el ámbito público debería garantizar la transparencia de la actividad algorítmica, en sus dos vertientes, de publicidad proactiva y del derecho de acceso, con el fin de que los ciudadanos puedan saber y comprender: (i) en qué ámbitos concretos de la actividad pública se adoptan decisiones basadas total o parcialmente en tratamientos algorítmicos; (ii) en qué medida esas decisiones algorítmicas afectan a los derechos y libertades de los ciudadanos, individual o colectivamente; (iii) cómo y bajo qué criterios se han adoptado esa clase de decisiones por parte de las instituciones públicas; (iv) y, por su incidencia en la racionalidad y eficiencia del gasto público, cómo se han adquirido e implementado esos sistemas de decisión algorítmica (e.g. mediante desarrollos *in house* o mediante licitación), qué fines institucionales específicos se pretenden cumplir, qué necesidades públicas, satisfacer y por qué dicha implementación tecnológica innovadora es la mejor alternativa frente a otras soluciones.

A la vista del estado del arte descrito hasta ahora, *de lege ferenda*, cualquier noción de transparencia aplicada al ámbito de la algoritmia decisional de las Administraciones públicas debería delimitarse con relación al quién (ámbito subjetivo), es decir, a quién es el sujeto destinatario de las obligaciones de transparencia (ámbito subjetivo); y al qué (ámbito objetivo), esto es, cuál es el contenido y alcance de las obligaciones de transparencia.

Desde el punto de vista subjetivo, la transparencia de los sistemas de IA debe considerarse respecto de los potenciales sujetos o grupos de interés a los que puede ir dirigida, por ejemplo, a la ciudadanía y conjunto de la sociedad civil, a través de determinadas obligaciones de publicidad activa o el ejercicio del derecho de acceso a la información pública; a los usuarios de estos sistemas (e.g. la Administración pública), importadores o distribuidores de los mismos; a los sujetos individuales afectados por una decisión adoptada o apoyada en estos modelos de aprendizaje automatizado (e.g. en el ámbito de un procedimiento administrativo concreto); al regulador y autoridades de control independientes que monitoricen la adecuación de los sistemas de IA a la legislación vigente; a expertos en análisis forense y auditores (internos o externos); a investigadores, expertos y comunidad científica.[239]

Desde el punto de vista del ámbito objetivo, cualquier regulación sectorial del uso de sistemas de toma de decisiones basados o mediante modelos ML debería concretar el contenido y alcance específico de las obligaciones de transparencia. En este sentido, el análisis de la literatura científica existente y de algunas propuestas elaboradas desde el ámbito institucional permiten identificar y sistematizar una serie de ítems o áreas comunes en los que la transparencia, en el sentido aquí explicado de interpretabilidad y explicabilidad, debería ser exigible en los sistemas de ML implementados por las organizaciones, incluyendo a las

Administraciones e instituciones públicas. Aunque no se trata de una lista exhaustiva, tales áreas comunes habitualmente identificadas son:[240]

(i) La responsabilidad sobre el modelo. Quién está involucrado en el desarrollo, implementación y explotación del modelo y a quién contactar, en su caso, para una revisión humana de las decisiones adoptadas por el sistema cuando éstas tengan impactos negativos en los derechos e intereses de los interesados.

(ii) Los fines, generales y concretos, del modelo. Cuando el modelo tenga múltiples fines, deberían determinarse cuáles son los prioritarios.

(iii) Los datos utilizados por modelo. Qué datos se han utilizado para entrenar el modelo, las fuentes de procedencia de esos datos, preparación de los datos (selección, ampliación, depuración y eliminación de datos redundantes o erróneos, combinación de datos de varias fuentes, determinación de variables relevantes, tratamiento de datos no estructurados, partición de los *datasets* en datos de entrenamiento, de prueba y validación del modelo), verificación de su calidad (fiabilidad de las fuentes de procedencia, representatividad de los datos, ausencia o minimización de los sesgos, actualización de los datos).

(iv) La identificación del modelo algorítmico utilizado. Los tipos de algoritmos seleccionados para generar el modelo a partir de los datos (e.g. modelos de regresión, árboles de decisión, bosques aleatorios, redes neuronales; justificación del modelo seleccionado en términos de rendimiento/interpretabilidad (en particular, si se trata de modelos de caja negra) frente a otros más interpretables u otras alternativas tecnológicas que permitan alcanzar los mismos fines establecidos por la organización; los parámetros del modelo, y los hiperparámetros seleccionados con carácter previo al entrenamiento del modelo; las métricas de rendimiento utilizadas para la validación del modelo y análisis de errores de entrenamiento y de generalización según el modelo utilizado.

(v) Fundamento y lógica del modelo. Las inferencias, patrones o correlaciones que justifican por qué unos datos de entrada generan unos resultados concretos, explicadas de forma inteligible, sencilla y no técnica; la identificación de las técnicas complementarias de interpretabilidad aplicadas para extraer explicaciones globales o locales, intrínsecas o *post-hoc* (e.g. empleo de contrafácticos para explicaciones individuales, modelos subrogados, gráfico de dependencias parciales, LIME o SHAP).

(vi) Impacto del modelo en términos de equidad. Identificación de las medidas adoptadas en el diseño e implementación del modelo para eliminar o mitigar los eventuales sesgos discriminatorios u otros efectos lesivos en los derechos e intereses individuales o colectivos, de manera que las decisiones adoptadas por el sistema sean imparciales, no discriminatorias y ajustadas a Derecho, garantizando el trato equitativo de los sujetos afectados por las decisiones del sistema.

(vii) Evaluaciones *ex ante* y *ex post* del modelo realizadas. Realización de evaluaciones de impacto (individuales y sociales) previas a la implementación y explotación del modelo; verificación del cumplimiento del desarrollador y del usuario del modelo con los requerimientos de transparencia algorítmica a través de auditorías internas y/o externas, que estaría a disposición de organismos reguladores y/o de control específicos y, en su caso, accesibles para el público en general mediante resúmenes ejecutivos.

(viii) Tales obligaciones de transparencia habrían de completarse con la exigencia de obligaciones de registro y documentación técnica que garanticen la transparencia e interpretabilidad del ciclo de vida completo de los modelos algorítmicos desarrollados e implementados por las Administraciones públicas y el sector público en general.

[1] Este artículo se ha realizado en el marco Grupo de Investigación "Good Governance for the Sustainable Development Goals" (GIGG_SDG) de la Universidad Rey Juan Carlos dirigido por Manuel Villoria en el marco del Proyecto Programa Interuniversitario en Cultura de la Legalidad (H2019/HUM-5699), ON TRUST-CM, coordinado por José María Saúca Cano (UC3M), del Programa de actividades de I+D entre grupos de investigación de la Comunidad de Madrid en Ciencias Sociales y Humanidades, cofinanciada con el Fondo Social Europeo.

El trabajo es la continuación de otras aportaciones realizadas anteriormente por la autora en el Grupo de Trabajo sobre Acceso a la Información Pública constituido dentro de la Red de Entidades locales por la Transparencia y participación ciudadana de la FEMP, coordinado por Joaquín Meseguer Yebra. La autora agradece al Grupo de Trabajo y a su coordinador la oportunidad única haber podido participar en distintas iniciativas y trabajo de campo desarrolladas por el Grupo en el ámbito del derecho a la información pública.

Asimismo. la autora agradece las interesantísimas observaciones realizadas por los revisores ciegos a los que se ha sometido la evaluación de este artículo, fundamentalmente, aquellas relacionadas con la aclaración de algunos conceptos técnicos a fin de hacerlos más divulgativos. Como resultado de este proceso de revisión se han incorporado nuevas referencias bibliográficas y ampliado las referencias casuísticas, y se ha añadido un nuevo epígrafe relativo a la transparencia algorítmica y a la explicabilidad que no constaba en el artículo original.

[2] This paper has been carried out within the framework of the Research Group "Good Governance for the Sustainable Development Goals" (GIGG_SDG) of Rey Juan Carlos University led by Manuel Villoria within the framework of the Interuniversity Program in Culture of Legality Project (H2019 / HUM- 5699), ON TRUST-CM, coordinated by José María Saúca Cano (UC3M), of R&D activities program among the Madrid Region research groups in Social Sciences and Humanities, co-funded by the European Social Fund.

The work is the continuation of other contributions previously made by the author in the Working Group on Access to Public Information set up within the Network of Local Entities for Transparency and Citizen Participation of the FEMP, coordinated by Joaquín Meseguer Yebra. The author thanks the Working Group and its coordinator for the unique opportunity to have been able to participate in different initiatives and field work developed by the Group in the field of the right to public information.

In addition. The author is grateful for the very interesting observations made by the blind reviewers to whom the evaluation of this article has been subjected, mainly those related to the clarification of some technical concepts in order to make them more informative. As a result of this review process, new bibliographic references have been incorporated and the casuistic references expanded, and a new section has been added regarding algorithmic transparency and explicability that did not appear in the original article.

[3] ICO (2017). *Big data, artificial intelligence, machine learning and data protection*, (September 4) v. 2.2.

[4] PASCAL, F. (2015). *The Black Box Society. The Secret Algorithms that Control Money and Information*, Cambridge (MA): Harvard University Press, pág. 5.

[5] UNITED NATIONS (2015*). Transforming our World. The 2030 Agenda for Sustainable Development*, pág. 28.

[6] COUNCIL OF EUROPE (2019). *Unboxing Artificial Intelligence: 10 Steps to Protect Human Rights*, Commissioner for Human Rights, págs. 9-10.

[7] DATENETHIKKOMMISSION (2019). *Gutachten der Datenethikkommission*, págs. 214-215.

[8] BELOT, L. (2016). «Amendement N°CL534». ASSEMBLEE NATIONALE, *République Numérique* (N° 3318), de 12 de enero.

[9] FINK, K. (2018). "Opening the government's black boxes: freedom of information and algorithmic accountability". *Information, Communication & Society*, vol. 21, núm. 10, pág. 1460. Al utilizar la investigación hasta cuatro fuentes distintas para realizar su estudio de campo, la autora advierte de que algunas de las solicitudes identificadas datan del año 2000.

[10] El número de resoluciones identificadas es a fecha de marzo 2021.

[11] STUCKE, M. E.; GRUNES, A. P. (2016). *Big Data and Competition Policy*, New York, Oxford University Press, págs. 16-28.

[12] AEPD (2020). *Tecnologías y Protección de Datos en las AA. PP.*, noviembre, pág. 30.

[13] COMISIÓN EUROPEA (2020). *Libro Blanco sobre la Inteligencia Artificial –un enfoque europeo orientado a la excelencia y la confianza*, COM(2020) 65 final, pág. 20.

[14] AI HLEG (2018). *A definition of AI: main Capabilities and Disciplines*. Brussels: European Commission, pág. 6. Disponible en https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines (consultado el 14 de marzo de 2021).

[15] COMISIÓN EUROPEA (2021). *Proposal for a Regulation of the European Parliament and of the Council laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts*. Brussels, 21.4.2021. COM(2021) 206 final.

[16] ICO (2017). *Big Data, Artificial Intelligence, Machine Learning and Data Protection*, versión 2.2, 4 de septiembre de 2017, págs. 6-7; COMISIÓN EUROPEA (2020). COM(2020) 65 final, *Op. cit*., pág. 20.

[17] THE ROYAL SOCIETY (2019). *Explainable AI: the Basics. Policy briefing*, pág. 6.

[18] En las tareas de "clasificación" el algoritmo predice una categoría o atributo a partir de un número predeterminado y finito de categorías. Por ejemplo, "este tumor es maligno o benigno" (2 categorías); "este contribuyente tiene una probabilidad alta, moderada o baja de defraudar a Hacienda" (3 categorías). En las tareas de "regresión", el algoritmo predice un valor numérico

(e.g. precio de venta de un inmueble; tiempo estimado de llegada de un vehículo a un destino concreto) en función de un número infinito de valores reales conocidos. Las tareas de "clustering" tratan de descubrir cuál es el mejor agrupamiento de los datos en función de factores o atributos comunes que no son evidentes para el análisis humano y consisten en dividir o segmentar el conjunto de datos de tal forma que los registros incluidos en uno de ellos se caracterizan por su gran similitud o cercanía. El "clustering" se utiliza, por ejemplo, para segmentar grupos de consumidores o usuarios por atributos (e.g. capacidad adquisitiva, lealtad a una marca) a fin de ofrecerles aquellos productos o servicios concretos en los que puedan estar más interesados; para la agrupación de acciones de un determinado índice bursátil (e.g. Ibex 35) en función de su comportamiento similar en un determinado periodo; para la compresión de imágenes o para la detección de comportamientos anómalos. Finalmente, en las tareas de "asociación", el algoritmo identifica patrones o relaciones no explícitas basadas en la ocurrencia conjunta de determinados grupos de atributos de ese conjunto de datos, de manera que esos patrones o relaciones dentro del *dataset* no sean el resultado de meras coincidencias no generalizables. CESEDEN (2013). *Big data en los entornos de Defensa y Seguridad*. Documento de investigación 03/2013, Instituto Español de Estudios Estratégicos, págs. 23-24; MARTÍNEZ HERAS, J. (2020). "¿Clasificación o Regresión?". *Guía rápida IArtificial.net*. Disponible en https://www.iartificial.net/clasificacion-o-regresion/ (consultado el 14 de marzo de 2021).

[19] Aunque también deben tenerse en cuenta el aprendizaje semi-supervisado y el aprendizaje por refuerzo. En el primer caso, a partir de un conjunto de datos etiquetados, se implementan técnicas de aprendizaje no supervisado para etiquetar otros datos de forma masiva. Es decir, se instruye al modelo para que aprenda utilizando aprendizaje supervisado de los datos históricos cuyos resultados son conocidos; y, después, se utiliza el modelo aprendido para etiquetar automáticamente el resto de los datos cuyos resultados se desconocen. Esta clase de aprendizaje se utiliza para la detección automática del *spam* o de anomalías y en el reconocimiento facial. En el "aprendizaje por refuerzo", el modelo realiza acciones y el aprendizaje se realiza a partir del refuerzo positivo (éxitos) o negativo (fracasos) recibido. La idea de que no todas las acciones tienen una recompensa, pero que toda acción tiene un valor que puede conducir posteriormente a una recompensa es esencial en el aprendizaje por refuerzo. Si el resultado de una decisión es beneficioso, el modelo aprende automáticamente a repetir esa acción en el futuro, desarrollando así estrategias a largo plazo que maximicen los beneficios. Esta clase de algoritmos se utilizan en el desarrollo de videojuegos, sistemas de navegación de drones y de coches autónomos, o en gestión de recursos humanos (e.g. turnos de empleados). DOMINGOS, P. (2018), *The Master Algorithm*…, *Op. cit*., págs. 219-222; MARTÍNEZ HERAS, J. (2020). "¿Cómo aprende la Inteligencia Artificial?", *Guía rápida IArtificial.net*, sp.

[20] Así, por ejemplo, los contenidos que sabemos gustan a un usuario (datos de entrada etiquetados) se utilizarían para predecir otros posibles contenidos que puedan gustarle en un futuro (datos de salida); la cotización bursátil de todos los lunes y martes durante los 10 últimos años, para predecir las cotizaciones de los martes en el futuro. Asimismo, en una tarea típica de clasificación, como sería la identificación de correos legítimos y *spam*, el destino sería una etiqueta que indica si un mensaje de correo electrónico es *spam* o no; y las variables podrían ser el remitente del correo electrónico, el texto del cuerpo del *email,* el texto en la línea de asunto, la hora de envío del mensaje de correo y la existencia de correspondencia anterior entre el remitente y el receptor, etc. AMAZON (sf.). *Amazon Machine Learning. Guía para Desarrolladores*, AWS, sp., pág. 10; MARTÍNEZ HERAS, J. (2020). "¿Cómo aprende…", *Op. cit*., sp.

[21] TRASK, A. W (2019). *Grokking Deep Learning*, New York, Manning Publications, págs. 11-13.

[22] *Idem*, págs. 22-34; ICO, *Big data…*, *Op. cit.,* pág. 10; VILLANUEVA, J. D. (2020). "Redes neuronales desde cero (I). Introducción". *Guía rápida IArtificial.net.*, *Op. cit*., sp. BASOGAIN, X. (s.f). *Redes neuronales artificiales y sus aplicaciones*. Escuela Superior de Ingeniería de Bilbao. Disponible en

https://ocw.ehu.eus/pluginfile.php/40137/mod_resource/content/1/redes_neuro/contenidos/pdf/libro-del-curso.pdf (consultado el 14 de marzo de 2021).

[23] VAN ECK, M. (2017). "Algorithms in public administration", *Bestuurecht & AI | goed bestuur bij technologi* [blog], pág. 2. Disponible en
https://marliesvaneck.wordpress.com/2017/01/31/algorithms-in-public-administration/
(consultado el 14 de marzo de 2021).

[24] CESEDEN (2013). *Big Data en los Entornos de Defensa y Seguridad*. Documento de investigación 03/2013. Instituto Español de Estudios Estratégicos, págs. 51-55; BABUTA, A.; OSWALD, M.; RINIK, C. (2018). *Machine Learning Algorithms and Police Secision-Making. Legal, Ethical, and Regulatory Challenges*. RUSI Whitehal Report 3-18, Universidad de Winchester.

[25] COGLIANESE, C.; LEHR, D. (2017). "Regulating by Robot: Administrative Decision Making in the Machine-Learning Era". *Georgetown Law Journal*, vol. 105, núm. 5, pág. 1161; APDCAT (2020), *Intel·ligència Artificial. Decisions Automatitzades a Catalunya*, Barcelona (enero), pág. 43.

[26] NEBRO MELLADO, J. J.; DE TORO MORÓN, A.; GARCÍA GONZÁLEZ, A., *et al*.(2020). "Diseño de algoritmos para la integración de la red social Twitter en el servicio público de limpieza de una Smart City", *IV Congreso de Ciudades Inteligentes*, Madrid, 15 de septiembre, sp.

[27] MANCOSU, G. (2019). « »Le contentieux des actes pris sur la base d'algorithmes, un point de vue italien ». *Revue Générale du Droit on line*, numéro 49010, pág. 2.

[28] O'NEIL, C. (2016). *Weapons of Math Destruction. How Big Data increases Inequality and threatens Democracy*. New YorK: Crown, págs. 14-18.

[29] Véase, ejemplo, la plataforma nacional francesa "Parcoursup", para la pre-inscripción de los estudiantes de secundaria en el primer año de enseñanza superior. MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR, DE LA RECHERCHE ET DE L'INNOVATION (sf.) *Parcoursup, c'est quoi?*, sp. Disponible en https://www.parcoursup.fr/index.php?desc=cest_quoi (consultado el 14 de marzo de 2021). Sobre el acceso al código fuente y al algoritmo de la aplicación informática Parcoursup, se ha pronunciado la Autoridad francesa de transparencia en diversas ocasiones. Vid. CADA, Opiniones 20201743, de 10 de septiembre 2020; 20184400, de 10 de enero de 2019; 2018245, 20182120 y 20182093 de 6 de septiembre de 2018; 20161990 de 23 junio de 2016, entre otras muchas.

[30] O'NEIL, *Weapons…*, *Op. cit*., págs. 168-169.

[31] APDCAT (2020). *Intel·ligència Artificial…*, *Op. cit*., págs. 34-41.

[32] SECRETARÍA DE ESTADO PARA LA SOCIEDAD DE LA INFORMACIÓN Y LA AGENDA DIGITAL (2017). *Plan Nacional de Territorios Inteligentes*, pág. 34; RAMIÓ, C. (2018), *Inteligencia artificial*

y Administración pública. Robots y humanos compartiendo el servicio público. Madrid: Catarata, pág. 14.

[33] *Natural Resources Defense Council* v. U.S. EPA, 954 F.3d 150 (April 1, 2020).

[34] PARLAMENTO EUROPEO (2020). *Resolución del Parlamento Europeo, de 20 de octubre, con recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas* (2020/2012(INL)).

[35] AUBY, J.-B. (2018). « Algorithmes et Smart Cities: Données Juridiques ». *Revue Générale du Droit*, núm. 29878, págs.15-18.

[36] CAPDEFERRO, Ó. (2019). "Las Herramientas Inteligentes Anticorrupción: entre la Aventura Tecnológica y el Orden Jurídico». *Revista General de Derecho Administrativo*, nº 50, pág. 6.

[37] Expediente 123/15-SV.

[38] Expediente CNMY18/AVSRE/4.

[39] Expediente 2020/LIC/0026.

[40] APDCAT (2020). *Intel·ligència Artificial. Decisions Automatitzades a Catalunya*, enero, Barcelona.

[41] COTINO HUESO, L. (2020). "SyRI, ¿a quién sanciono? Garantías frente al uso de inteligencia artificial y decisiones automatizadas en el sector público y la sentencia holandesa de febrero de 2020". *LA LEY privacidad*, núm. 4, Wolters Kluwer, LA LEY 4999/2020.

[42] KOENE, A.; CLIFTON, C.; HATADA, Y., *et al*. (2019). "A governance framework for algorithmic accountability and transparency", European Parliamentary Research Service, Scientific Foresight Unit (STOA), pág. 56.

[43] MANCONSU, G. (2019). « Les algorithmes publics déterministes au prisme du cas italien de la mobilité des enseignants ». *Rivista Italiana di Informatica e Diritto*, núm. 1, pág. 76.

[44] DEPARTMENT FOR BUSINESS, ENERGY & INDUSTRIAL STRATEGY; DEPARTMENT FOR DIGITAL, CULTURE, MEDIA & SPORT; OFFICE FOR ARTIFICIAL INTELLIGENCE (2020). *Guidelines for AI procurement*, sp.

[45] CERRILLO I MARTÍNEZ, A. (2020). "¿Son fiables las decisiones de las Administraciones Públicas adoptadas por algoritmos?", *European Review of Digital Administration & Law*, Vol. 1, Issue 1-2, Erdal, pág. 22.

[46] Para COTINO, *op. cit.*, *sería una obligación legal difundir estos tratamientos en razón del artículo 6 bis de la Ley 19/2013, de 9 de diciembre, de transparencia, respecto del inventario de actividades de tratamiento en aplicación del artículo 31 de la Ley orgánica 3/2018*. Sin embargo, debe apuntarse que esta obligación de incluir en el Registro de Actividades los tratamientos algorítmicos se limitaría exclusivamente a los tratamientos que utilizasen datos personales, lo cual seguiría dejando en la "opacidad" tratamientos algorítmicos con datos no personales que, por ejemplo, tuvieran impacto en el diseño de políticas públicas, y por tanto, en la vida de los

ciudadanos, en sencillamente decisiones automatizadas donde el interesado sea una persona jurídica.

[47] Cfr. AI HLEG (2019). *Directrices Éticas para una IA Fiable*. Bruselas: Comisión Europea, pág. 20; VIDA FERNÁNDEZ, J. (2018). "Los retos de la regulación de la inteligencia artificial: algunas aportaciones desde la perspectiva europea", en: DE LA QUADRA-SALCEDO, T.; PIÑAR MAÑAS, J. L. (Dirs.) *Sociedad Digital y Derecho*. Madrid: Boletín Oficial del Estado, Ministerio de Industria, Comercio y Turismo y RED.ES, pág. 218.

[48] PARLAMENTO EUROPEO, 2020/2012(INL), *Op. cit*., apartado N.

[49] COGLIANESE, C.; LEHR, D. (2017). "Regulating by Robot…", *Op. cit*., pág. 1205.

[50] DESAI, D. R.; KROLL, J. A. (2017). "Trust but verify: A guide to algorithms and the law". *Harvard Journal of Law & Technology*, 31 (1), págs. 8, 13-14.

[51] Véanse, entre otros, *EPIC v DHS* - Joint Status Report 15-cv-00289 May 28 2015-FINAL; DHS (2011). *Privacy Impact Assessment Update for the Future Attribute Screening Technology (FAST)/Passive Methods for Precision Behavioral Screening*, DHS/S&T/PIA-012(a), 21 de diciembre. De los documentos facilitados y publicados online por el DHS se supo que el programa FAST había sido testado en un "lugar amplio" del "nordeste" del país, y aunque no se determinó el emplazamiento, sí que se excluyó que fuese un aeropuerto. También se conoció que, a través de una red sensores "no intrusivos" se obtenía información personal de individuos en el curso de su actividad diaria procedente de imágenes de vídeo, grabaciones de audio, señales cardiovasculares, feromonas, actividad electro-dérmica y medición respiratoria; y que tal información era procesada por un algoritmo cuya finalidad era detectar "anormalidades comportamentales" clasificadas como "comportamiento engañoso" o "mal-intencionado".

[52] Cfr. EPIC v DHS FAST Complaint FINAL, Case 1:15-cv-00289-CKK, February 26, 2015, para 7 [solicitud de medidas cautelares al Tribunal del Distrito de Columbia].

[53] La regresión logística es un algoritmo de aprendizaje supervisado que clasifica binariamente en función de la probabilidad de que un resultado sea "1ó 0", "sí o no", "verdadero o falso", minimizando el error entre el valor predicho y el de los registros del *dataset* que forman parte del conjunto de entrenamiento. Vid. CESEDEN (2013), *Op. cit.*, pág. 25.

[54] ICO. FS50798023, de 29 de enero.

[55] CTBG, R/0051/2017, de 25 de abril. Antecedente de hecho núm. 4. La automatización comprende desde que un cinemómetro detecta una posible infracción por exceso de velocidad, captura, compacta, encapsula y encripta mediante un algoritmo *hash* las imágenes digitales y el fichero de texto con la información referente a la infracción; después se remiten al CTDA, donde nuevamente se tratan los ficheros y se cotejan matrícula y modelo de vehículo de la imagen con el que consta en el Registro de Vehículos; hasta que finalmente se genera el correspondiente expediente sancionador en la DGT.

[56] Los requisitos para ser considerado "consumidor vulnerable" y ser beneficiario del bono social están comprendidos en el art. 3 del Real Decreto 897/2017, de 6 de octubre, por el que se regula

la figura del consumidor vulnerable, el bono social y otras medidas de protección para los consumidores domésticos de energía eléctrica; y desarrollados por la Orden ETU/943/2017, de 6 de octubre.

[57] CTBG. Resolución 701/2018, de 18 de febrero de 2019. Antecedentes de hecho 1 y 3.

[58] *Manifiesto en favor de la transparencia en desarrollos de software públicos*. Disponible en https://transparenciagov2020.github.io/#manifiesto (consultado el 14 de marzo de 2021). Entre la documentación técnica solicitada se encontraban el repositorio con el código y la versión de todos los elementos del sistema en el momento de su publicación así como futuros cambios, incluyendo aplicación y servidores junto con los detalles de su despliegue y gobernanza: dónde se hallan, quién los administra, y qué medidas de seguridad se adoptaron tanto para el despliegue a nivel nacional como a nivel autonómico; el repositorio de código utilizado durante su desarrollo, su historial desde el inicio del desarrollo y los cambios desde la primera versión disponible; el informe de diseño del sistema (aplicación y servidores), detallando los análisis que han llevado a decidir los parámetros de configuración y uso de la API de Exposición de Notificaciones, las librerías y servicios utilizados para evaluar la seguridad y privacidad de los datos; el informe detallado de los mecanismos de monitorización de la aplicación y los asociados para asegurar la privacidad y el cumplimiento de la normativa de protección de datos, así como la evaluación de impacto en la protección de datos de la aplicación asociada a su uso en las plataformas Android y iOS.

[59] BOURCIER, D.; DE FILIPPI, P. (2018). . "Les algorithmes sont-ils devenus le langage ordinaire de l'administration ? ". KOUBI, G.; CLUZEL-METAYER, L.; TAMZINI, W. (2018). *Lectures critiques du Code des relations Public et administration*, LGDJ, págs.193-210.

[60] GAIP. Resolución de 21 de septiembre de 2016, estimando las Reclamaciones 123/2016 y 124/2016 (acumuladas), F.J.3.

[61] SENTENCIA DEL TRIBUNAL GENERAL (Sala Tercera) de 19 de marzo de 2010, T-50/05, Evropaïki Dynamiki/Comisión, apartado 81.

[62] CONCLUSIONES DEL ABOGADO GENERAL SR. PAOLO MENGOZZI, presentadas el 21 de marzo de 2013, C-657/11, Belgian Electronic Sorting Technology, nota 7.

[63] Para que el código fuente sea ejecutable debe ser compilado en "código máquina" o "código objeto" a través de un programa denominado "compilador", mediante el cual es convertido en instrucciones expresadas en lenguaje binario legibles por la máquina e ininteligibles para los humanos. Habitualmente, en el modelo de licencia propietaria característico de los sistemas de propiedad intelectual, el software se facilita en forma de código objeto que es ejecutable pero difícilmente modificable, pues la conversión (inversa) del código objeto al código fuente suele ser muy compleja técnicamente. Precisamente, el movimiento *Free and Open Software* ("FOSS") defiende la inversión de este modelo propietario de la industria tradicional, facilitando el código fuente original al usuario lo que simplifica el desarrollo posterior del software, mediante su modificación o la inclusión de nuevas líneas de código. Cfr. WALDEN, I. (2013). "Open Source as philosophy, methodology and commerce". En SHEMTOV, N. & WALDEN, I. *Free and Open Software. Policy, Law and Practice*. Oxford: Oxford University Press, págs. 1-2; RUSTAD, M. L. (2010). *Software Licensing. Principles and Practical Strategies*. Oxford: Oxford University Press, págs. 101-102.

[64] SENTENCIA DEL TRIBUNAL DE JUSTICIA (Sala Tercera), de 22 de diciembre de 2010, Bezpečnostní softwarová asociace, C-393/09, apartado 34.

[65] Opiniones núms. 20144578, de 8 de enero de 2015; 20161990, de 23 de julio de 2016; 20161989, de 23 de julio de 2016; 20180376, de 31 de mayo de 2018; 20182682 de 6 de septiembre de 2018; 20182455, de 6 de septiembre de 2018.

[66] GAIP. Resolución de 21 de septiembre de 2016, estimando las Reclamaciones 123/2016 y 124/2016 (acumuladas), F.J.3, citada *supra*.

[67] CTBG. R/0701/2018, de 18 de febrero de 2019, FD 5.

[68] Clear One Commc'ns, Inc. v. Chiang, 608 F. Supp. 2d 1270, 1273 (D. Utah 2009); Morley v. Square, Inc., No. 4:10CV2243 SNLJ, 2016 WL 1615676, at *2 (E.D. Mo. Apr. 22, 2016).

[69] GAIP. Resolución de 21 de septiembre de 2016, estimando las Reclamaciones 123/2016 y 124/2016 (acumuladas), F.J.3, citada *supra*.

[70] OSWALD, M.; GRACE, J.; URWIN, S.; BARNES, G. C. (2018). "Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality", *Information & Communications Technology Law*, Vol. 27, núm. 2, págs. 223-228.

[71] DOMINGOS, P. (2018). *The Master Algorithm…*, *Op. cit*., pág. 24.

[72] Cfr. ICO; ALAN TURING INSTITUTE (2020). *Op. cit.*, págs. 50-ss.

[73] COGLIANESE y LEHR (2017). *Op. cit*., págs. 1157-1158.

[74] Cfr. DOMINGOS, P. (2018). *The Master Algorithm…*, *Op. cit*., pág. 237- 239.

[75] MANCOSU (2019). "Les algorithmes publics déterministes…" *Op. cit*., pág. 75.

[76] AUBY (2018). "Algorithmes et Smart Cities…". *Op. cit*., pág. 21.

[77] Cfr. CTBG. R/0701/2018, de 18 de febrero. Antecedentes de Hecho núm. 3, donde en el trámite de alegaciones frente al Consejo, la Subdirección General de Tecnologías de la Información y las Comunicaciones considera que, con relación al código fuente del bono social *cabría la inadmisión de la solicitud, ya que este código no se considera información pública según el artículo 13 de la Ley de transparencia al no ser ni "contenidos" ni "documentos", sino programas informáticos*.

[78] Cfr. DEPARTMENT OF JUSTICE. *Report on Electronic Record FOIA Issues, Part II. FOIA Update*, vol. XI, núm. 3, págs. 3-12.

[79] MANCOSU (2019). "Les algorithmes publics déterministes…" *Op. cit.*, pág. 77.

[80] *Gilmore v. US Dept. of Energy*, 4 F. Supp. 2d 922 (N.D. Cal. 1998). El Tribunal del Distrito Norte de California analiza una resolución desestimatoria de acceso al amparo de la FOIA con relación a

> *CLEVERER un software de videoconferencia desarrollado por un contratista del Departamento de Energía (DOE) y sujeto a una licencia no exclusiva de uso. En su sentencia, el Tribunal concluyó que CLEVERER no podría considerarse información de la agencia incluso si el DOE fuese titular originario o controlase [el software], porque no facilita información sobre el funcionamiento, estructura o procesos de toma de decisiones de la Administración*.

[81] ICO, IC-43073-B3W7, de 16 de diciembre 2020, paras. 17-25. Al parecer, detrás de la solicitud estaba la circunstancia de que se habían hecho públicos algunos problemas de seguridad de la aplicación relacionados con la inadecuada protección de la confidencialidad de las comunicaciones. De hecho, el propio solicitante publicó en el sitio web "WhatDoTheyKnow.com" un enlace a una nota de prensa de la Comisión Federal de Comercio de los EE.UU. (FTC), donde la Agencia señalaba que, desde el año 2016, la plataforma había inducido a error de forma sistemática a sus usuarios sobre el nivel de seguridad real de la aplicación. En concreto, la FTC consideraba que Zoom había venido ofreciendo "encriptación de 256 bits, de punto a punto", cuando en la práctica el nivel de seguridad era menor; había conservado grabaciones sin encriptar durante más de 60 días en sus servidores antes de transferirlos a una nube segura; y, había guardado, además, las claves criptográficas que le permitirían acceder al contenido de las reuniones virtuales. Vid. *FTC Requires Zoom to Enhance its Security Practices as Part of Settlement*, 9 de noviembre de 2020. Disponible en https://www.ftc.gov/news-events/press-releases/2020/11/ftc-requires-zoom-enhance-its-security-practices-part-settlement (consultado el 14 de marzo de 2021).

[82] LOI n° 2016-1321 du 7 octobre 2016 pour une République Numérique. JORF n° 0235 du 8 octobre 2016.

[83] CADA. Opinión núm. 20142953, de 16 de octubre de 2014, donde se estimó el acceso a un programa informático desarrollado por una empresa privada, relacionado con un contrato público para la construcción del Musée des Confluences de Lyon, con eliminación de aquellas partes que pudieran verse afectadas por el secreto en materia industrial o comercial; Opinión núm. 20144578, de 8 de enero de 2015, en la que se estimó el derecho de acceso al código fuente de un software elaborado por la Dirección General de Finanzas que simulaba el cálculo del impuesto sobre la renta, para poder reutilizarlo en un trabajo de investigación universitario, pues se consideró que los archivos informáticos que constituían el código fuente solicitado, producido por la Administración como parte de su misión de servicio público, tenían el carácter de documentos administrativo; Opinión 20161990, del 23 de junio de 2016, en la que se consideró que el algoritmo desarrollado por el Ministerio de Educación francés, conocido como «APB» (*Admission Post-Bac*) para la tramitación de las solicitudes de admisión a los grados universitarios tenía también la consideración documento administrativo en el sentido del CRPA, y por tanto, era accesible.

[84] Con relación a resoluciones estimatorias del derecho de acceso al código fuente, véanse, CADA. Opiniones 20191797, de 16 de enero de 2020; 20182093, de 6 de septiembre de 2018;

20182120, de 6 de septiembre de 2018; 20182455, de 6 de septiembre de 2018; 20182682, de 6 de septiembre de 2018; 20180276, de 19 de abril de 2018; 20161990, de 23 de junio de 2016; 20161989, de 23 de junio de 2016. En sentido desestimatorio, véanse también las Opiniones núms. 20200496, de 12 de marzo de 2020; 20181891, de 18 de julio de 2019; 20184400, de 10 de enero de 2019; 20180376, 31 de mayo de 2018.

[85] En sentido estimatorio, CADA. Opiniones núms. 20182093, de 6 de septiembre de 2018; 20182120, de 6 de septiembre de 2018; 20182455, de 6 de septiembre de 2018; 20173235, de 30 de noviembre de 2017; 20163835, de 6 de octubre de 2016; 20161990, de 23 de junio de 2016; 20161989, de 23 de junio de 2016; en sentido desestimatorio, 20201743, de 10 de septiembre de 2020; 20184400, de 10 de enero de 2019.

[86] CADA. Opiniones núms. 20184400, de 10 de enero de 2019; 20182093, de 06 de septiembre de 2018; 20182120 y 20182455, de 6 de septiembre de 2018, respectivamente.

[87] CADA. Opinión núm. 20182093, de 06 de septiembre de 2018. De hecho, el Ministerio de Enseñanza Superior y de Investigación francés (MESRI) viene publicando en la plataforma de software colaborativo «Gitlab» (https://framagit.org/jrfaller/algorithmes-de-parcoursup) las distintas actualizaciones de los algoritmos y código Java que permiten el cálculo del orden de llamada, las propuestas de formación, las propuestas de alojamiento en residencias universitarias, la aplicación del esquema “Mejores bachilleres”, y la aplicación del contestador automático.

[88] CADA. Opinión núm. 20181891, de 18 de julio de 2019.

[89] GAIP. Resolución 93/2019, de 22 de febrero, FJ 3.

[90] GAIP. Resolución 200/2017, de 21 de junio, F.J.2. La solicitud tiene por objeto el acceso al código fuente de la aplicación empleado por la Administración para la designación de los miembros de los tribunales correctores de las pruebas de acceso a la universidad (PAU) y que condiciona el orden en el listado de miembros correctores de las PAU mediante una relación inequívoca entre el valor numérico del número aleatorio extraído en el sorteo, el número del solicitante y el número de la posición en la lista resultante, con la finalidad de comprobar que el programa está correctamente diseñado y que la selección de los miembros de los tribunales se hace realmente por sorteo.

[91] CITRON, D. (2008). “Technological Due Process”. *Washington University Law Review*. núm. 85, págs. 1254-1255; BOURCIER y DE FILIPPI (2018). “Les algorithmes sont-ils devenus…”. *Op. cit*. págs. 200 y 207.

[92] CITRON, D.; CALO, R. (2019). *The Automated Administrative State*. Harvard Kennedy School, Shorenstein Center, sp.

[93] BOURCIER, D.; DE FILIPPI, P. (2018). “La transparence des algorithmes face à l’Open Data: Quel statut pour les données d’apprentissage? ” En: *Revue française d’Administration Publique*, ENA, págs. 7, 12-14, pág. 15.

[94] BINNS, R.: GALLO, V. (2019). *Automated Decision Making: the Role of Meaningful Human Reviews*, Information Commissioner's Office, sp.; VEALE, M.; VAN KLEEK, M.; BINNS, R. (2018). "Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making". *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April. Paper No. 440, pág. 4.

[95] BOURCIER y DE FILIPII (2018). "Les algorithmes sont-ils devenus...", *Op. cit.*, pág. 207; OSWALD, M. (2018). "Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power". *Philosophical Transactions of the Royal Society A.*, sp.

[96] AUBY (2018), "Algorithmes et Smart Cities…". *Op. cit.*, pág. 21.

[97] BOURCIER, D.; DE FILIPPI, P. (2018). "Les algorithmes sont-ils devenus…". *Op. cit*., pág.195.

[98] Cfr. SCHARTUM (2016), "Law and algorithms in the public domain", *Etikk i praksis. Nordic Journal of Applied Ethics*, Núm 1, págs. 16-18.

[99] DE LA QUADRA-SALCEDO, T. (2018). "Retos, riesgos y oportunidades de la sociedad digital". En: DE LA QUADRA-SALCEDO; PIÑAR MAÑAS, *Op. cit*. pág. 54.

[100] *K.W. v. Armstrong*, 180 F. Supp. 3d 712-715 (D. Idaho 2016). A partir de las declaraciones realizadas por analistas y del expediente remitido por el IDHW, se concluyó que de la muestra original de datos clínicos procedentes de 3.512 pacientes utilizados para construir el modelo, al menos un 66% tuvieron que ser excluidos posteriormente por existencia de numerosos errores en los datos de la muestra; determinadas áreas de población estaban infrarrepresentada estadísticamente con relación a su tamaño; la herramienta asignaba una prestación inferior a la que realmente correspondía a al menos un 15% de los interesados. Errores en el diseño del modelo que, junto a la ausencia de un procedimiento de auditoría, determinaban la ausencia de fiabilidad de la herramienta.

[101] ALAMILLO DOMINGO, I. y URIOS APARISI, X. (2011). *La actuación administrativa automatizada en el ámbito de las Administraciones Públicas. Análisis jurídico y metodológico para la construcción y la explotación de trámites automáticos*. Barcelona: Escola d'Administració Pública de Catalunya, págs. 101-103.

[102] AI NOW INSTITUTE (2018). *Litigating algorithms: challenging government use of algorithmic decision systems*, pág. 8. Disponible en https://ainowinstitute.org/litigatingalgorithms.pdf (consultado el 14 de marzo de 2021); *Taking Algorithms To Court. Current Strategies for Litigating Government Use of Algorithmic Decision-making*. Disponible en https://ainowinstitute.org/announcements/litigating-algorithms.html (consultado el 14 de marzo de 2021).

[103] CTBG. R/0051/2017, de 25 de abril de 2017, FD 6. En el Antecedente núm. 5, el reclamante señala que la respuesta dada por la Administración no ha determinado cuál es el margen de error en el cinemómetro; cómo se procede a identificar una velocidad a partir de un dato registrado en un equipo de medida, cuya incertidumbre se desconoce o de la que no se informa; cómo se ha trasladado al expediente sancionador ese margen de error, ni tampoco *cómo se ha aplicado hasta llegar a la velocidad reflejada de 72 km/h, objeto de la sanción*".

[104] GAIP. Resolución 200/2017, de 21 de junio, F.J.2. En sentido similar, véase la Resolución de 21 de septiembre de 2016, de estimación de les Reclamaciones 123/2016 i 124/2016 (acumuladas), FJ 3, donde la Autoridad catalana considera inaplicable el límite de la confidencialidad de los procesos de toma de decisiones con relación al acceso solicitado del algoritmo:

> *Por otra parte, no se entiende cómo el acceso al algoritmo matemático que sirve de punto de partida del programa informático que implementa el proceso puede afectar negativamente el funcionamiento de éste. Como se ha visto, el resultado del procedimiento viene determinado por un sorteo entre todos los aspirantes que reúnan los requisitos mínimos establecidos, y en el que sólo hay que observar los requerimientos de paridad entre hombres y mujeres y de porcentajes mínimos de profesores universitarios y de bachillerato establecidos por el Real Decreto 1892/2008. El algoritmo, como se ha visto, se ha de limitar y se limita a recoger y aplicar estas variables (y las otras antes mencionadas), que son regladas y no confieren ningún margen de discrecionalidad, por lo que no se ve qué interés puede haber en ocultarlo y mantenerlo confidencial. En cambio, sí existe un interés público y privado -de interesados- evidente en poder comprobar que el algoritmo que guía todo el proceso está correctamente diseñado para garantizar la igualdad de todos los participantes en el proceso selectivo [subrayado nuestro].*

[105] Cfr. CTBG. Resolución 701/2018, de 18 de febrero de 2019, FFJJ 4, 5 y 6.

[106] CIVIO (2019). *Recurso contencioso-administrativo contra la Resolución 701/2018 del Consejo de Transparencia y Buen Gobierno de fecha 18 de febrero de 2019*, FD II. Primero. Disponible en https://civio.app.box.com/s/l23h4100guo9ol2f65a7kbcl1uglm30m (consultado el 14 de marzo de 2021).

[107] ICO (2017). *Big data, artificial intelligence,* Op. cit., pág. 10.

[108] DESAI y KROLL (2017). «Trust but verify…». *Op. cit*., pág. 3.

[109] BATHAEE, Y. (2018). "The Artificial Intelligence Black Box and the Failure of Intent and Causation". En *Harvard Journal of Law & Technology*, vol. 31(2) pág. 905. Véase también, PARALAMENTO EUROPEO (2020). Resolución 2020/2012(INL), Op. cit., apartado 23.

[110] ICO y ALAN TURING (2020), *Op. cit*. págs. 66, 115-117. En sentido similar, véase también DOMINGOS, P. (2018). *The Master Algorithm… Op. cit.*, pág. 238.

[111] LIU, H.-W.; LIN, C.-F.; CHEN, Y.-J. (2019). "Beyond State v Loomis: artificial intelligence, government algorithmization and accountability". En: *International Journal of Law and Information Technology*, Vol. 27 (2), págs. 135–136.

[112] AI HLEG (2019). *Ethics Guidelines for Trustworthy AI*. Brussels: European Commission, 8 de abril, pág. 21.

[113] DE LAAT, P. B. (2018). "Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?", en *Philosophy & Technology*, Núm. 31, pág. 537.

[114] ICO y ALAN TURING (2020). *Explaining decisions*…, *Op. cit*. pág. 70.

[115] ECLI: NL: RVS: 2017: 1259-Raad van State, de 17 de mayo de 2017, nr. 201600614/1/R2 y otros, 14.3 y 14.4.

[116] FINK, K. (2018), "Opening the government's black boxes". *Op. cit.*, pág. 1454.

[117] BOURCIER y DE FILIPPI (2018). "La transparence des algorithmes…" *Op. cit.*, pág. 7.

[118] DIAKOPOULOS, N. (2016). "Accountability in Algorithmic Decision Making". *Communications of the ACM*, vol. 59, núm. 2, págs. 57-58.

[119] HASSAN; S.; DE FILIPPI, P. (2017) "The Expansion of Algorithmic Governance: From Code is Law to Law is Code", *Field Actions Science Reports*, Special Issue 17, págs. 88-90.

[120] DE LA CUEVA, J. (2020). "Código fuente, algoritmos y fuentes del Derecho". *El Notario del Siglo XXI*, Núm. 89, ENERO – FEBRERO. Disponible en: http://www.elnotario.es/index.php/opinion/8382-codigo-fuente-algoritmos-y-fuentes-del-derecho (consultado el 14 de marzo de 2021).

[121] Cfr. ICO (2018). *ICO Investigation into how the Police use Facial Recognition Technology in Public Places*, 31 de octubre, pág. 34, donde la autoridad inglesa señala que las mujeres y las minorías étnicas son más sensibles a los falsos positivos.

[122] CITY OF PORTLAND, *Ordinance 190113. Prohibit the acquisition and use of Face Recognition Technologies by the City of Portland Bureaus*, apartados 9 y 12 (parte expositiva), y apartado f) (parte dispositiva). La Ordenanza en cuestión limita el uso de tales tecnologías exclusivamente al reconocimiento biométrico con fines de autenticación (one-to-one) de los empleados públicos para acceder a sus dispositivos electrónicos de trabajo, para la detección automática de rostros a fin de eliminarlos de las grabaciones de las sesiones municipales objeto de publicidad activa a fin de proteger la privacidad, así como la detección automática de rostros en las cuentas oficiales en redes sociales de las Autoridades municipales. En el caso de la detección automática de rostros en cuentas oficiales en redes sociales de las Autoridades municipales, debe tenerse en cuenta que, en las disposiciones específicas sobre esta clase de cuentas previstas en la Norma Administrativa de Recursos Humanos, *HRAR 4.08 (A). Social Media*, de 2011, se establece que, antes de publicar imágenes o videos en una cuenta oficial en redes sociales de las Autoridades Públicas deberá verificarse la protección de la eventual expectativa de privacidad de aquellas terceras personas que aparezcan en la imagen.

[123] ACLU (2020). *Man wrongfully arrested because face recognition can't tell black people apart*, Detroit, 20 de junio. Disponible en https://www.aclu.org/press-releases/man-wrongfully-arrested-because-face-recognition-cant-tell-black-people-apart (consultado el 14 de marzo de 2021).

[124] ACLU (2019). *ACLU Submits Freedom of Information Request Seeking All Records Regarding Detroit Police Department's Use of Facial Recognition Technology*, 17 de septiembre, Detroit. Disponible en https://www.aclumich.org/en/press-releases/civil-rights-coalition-urges-detroit-board-police-commissioners-reject-detroit-police (consultado el 14 de marzo de 2021). Entre la información solicitada se incluían, entre otra, memoranda, comunicaciones internas e interdepartamentales, también correos electrónicos, decisión de contratación, modificaciones contractuales, documentación descriptiva de la solución tecnológica y resultados del testado y validación del sistema para personas de color en comparación con personas de rasgos caucásicos.

[125] En ocasiones, puede resultar imposible indicar las razones que justifican la confidencialidad de una determinada información o documento, sin divulgar su contenido y, por lo tanto, sin privar a la excepción de su finalidad esencial, cuál es la protección de un bien jurídico específico. Es en tales casos que el Derecho anglosajón acude a la "doctrina Glomar" en EE.UU. o las "respuestas NCND" (*neither confirm, neither deny*) en el Reino Unido, y habitualmente, se aplica cuando la información solicitada pueda afectar a la seguridad y la defensa nacional, a las relaciones internacionales, o a la detección, prevención, investigación y enjuiciamiento de ilícitos penales y la protección de la seguridad pública. En alguna ocasión, esta doctrina también ha sido aplicada en el ámbito del Derecho Europeo, al amparo del art. 9.3 del Reglamento (CE) n° 1049/2001 del Parlamento Europeo y del Consejo, de 30 de mayo de 2001, relativo al acceso del público a los documentos del Parlamento Europeo, del Consejo y de la Comisión. Cfr. STJCE. Sisón/Consejo, T-110/03, T-150/03 y T-405/03, de 26 de abril de 2005, apartado 60. STJCE; confirmado por Sisón/Consejo, C-266/05 P, de 1 de febrero de 2007, apartados 101, con relación a la protección de la seguridad pública y las relaciones internacionales.

[126] Esta excepción limitaría el derecho de acceso a aquella información relativa a técnicas, procedimientos y directrices de investigación y enjuiciamiento de ilícitos penales y administrativos –e incluso de naturaleza civil– por las Autoridades competentes (e.g. Administraciones con competencias en materia de seguridad pública, Fuerzas y Cuerpos de Seguridad) cuando razonablemente existe un riesgo de que tal divulgación pueda ser utilizada para eludir la aplicación de la Ley. La invocación de la "doctrina Glomar" ha sido admitida por los Tribunales Federales cuando la mera divulgación del uso de una técnica de investigación concreta revele las circunstancias bajo las cuales dicha técnica se ha utilizado. THE UNITED STATES DEPARTAMENT OF JUSTICE (2019). *Department of Justice Guide to the Freedom of Information Act. Exemption 7*, 14 de mayo, págs. 1, 10-11.

[127] *Am. Civil Liberties Union Found. v. DOJ*, 2019 BL 443104, N.D. Cal., No. 19-cv-00290-EMC:

> *[…] el riesgo de que la actividad delictiva escape de la detección a través de redes sociales si el FBI reconoce que no dispone de esa información queda sustancialmente mitigado por dos hechos. Primero, es bien conocido que las otras agencias relacionadas están realizando actividades de vigilancia en redes*

> *sociales en los centros de inmigración y están compartiendo esa información. Ello aminora el riesgo de que, ante la respuesta del FBI, haya quienes se atrevan a difundir mensajes criminales o terroristas a través de las redes sociales. Incluso si el FBI reconociese que no ha comprado o adquirido productos y servicios para monitorizar las redes sociales, ello no significa que el FBI no tenga tales herramientas a su disposición porque podría haberlas desarrollado internamente.*

[128] Véase, por ejemplo, MINISTERIO DE JUSTICIA. *Expediente 001-005871, de 20 de abril de 2016*. La solicitud tuvo por objeto el *acceso al código fuente de la aplicación informática LEXNET […] para comprobar con ayuda de un perito que se encuentra bien desarrollada técnicamente y no genera problemas de seguridad*. Resulta discutible, sin embargo, la fundamentación jurídica de la resolución denegatoria con base en la aplicación del límite en el derecho de propiedad industrial (art. 14.1.j de la LTAIBG), al considerar que LEXNET es *una marca registrada por el Ministerio de Justicia en el Registro de Propiedad Industrial*. Como acertadamente ha señalado Esteve Pardo, el objeto de la solicitud no era la utilización de la marca LEXNET para comercializar programas de ordenador, sino el acceso al código fuente para verificar su correcto funcionamiento, dado el gran número de incidencias que esta aplicación había registrado. Vid. ESTEVE PARDO, M. A. "El secreto profesional y la propiedad intelectual e industrial". ARAGUAS GALERA, I.; PÉREZ GARCÍA, I. L. *et. al*. (2017)). *Los límites al derecho de acceso a la información pública*. Madrid: INAP, pág. 182. Aunque la resolución desestimatoria fue impugnada ante el CTBG, sin embargo, el recurso fue objeto de archivo al no subsanarse en plazo por el reclamante las deficiencias en la presentación de la reclamación ante el Consejo durante el trámite correspondiente. Vid. CTBG. R/0275/2016, de 22 de agosto, FJ 3.

[129] *Elkins v. FAA*, No. 14-1791, 2015 WL 2207076 (D.D.C. May 12, 2015) (Boasberg, J.), con relación a una solicitud de acceso dirigida a la Administración Federal de Aviación (FAA) que tenía por objeto conocer las circunstancias por las cuales un avión no identificado sobrevoló en círculos el domicilio del solicitante en una fecha y hora concreta. Entre la información solicitada se encontraban las comunicaciones entre el Departamento de Justicia y la FAA relativas al vuelo en cuestión, los requerimientos realizados a la FAA relativos a posibles diligencias de investigación abiertas con relación al solicitante, el trayecto de vuelo en el radar, el número de registro del aparato (N Number) y la información relativa al Código MS del transpondedor vinculado a la aeronave.

[130] En concreto, la Sección 12(1) de la *Freedom of Information Act* de 2000 contempla la posibilidad de inadmitir una solicitud de acceso a la información pública cuando el coste de tramitación estimado exceda del límite reglamentariamente establecido en las *Fees Regulations 2004*. Para el Gobierno central, las cámaras legislativas y las Fuerzas Armadas dicho límite del coste de tramitación sería de 600 libras (680.81 €) y, para el resto de Administraciones y entidades públicas, dicho límite se correspondería con 450 libras (511.57 €). Nótese que este límite específico de la normativa inglesa no tiene equivalente en las causas de inadmisión previstas en el art. 18.1 LTAIBG y normativa autonómica concordante.

[131] ICO. FS50798023, de 29 de enero, paras. 36-47. El ICO estimó que el límite del coste de la tramitación de la solicitud para la Policía de Norfolk estaría fijado en un total de 18 horas, cuyo coste económico se correspondería con el límite reglamentario de 450 libras, es decir, una media de 25 libras/hora.

[132] FINK (2018). "Opening the government's black boxes…". *Op. cit*., pág. 1455.

[133] MINISTÈRE DE L'INTERIEUR (2016). *Lancement de l'application mobile SAIP : Système d'alerte et d'information des populations*. Délégation à l'Information et à la Communication du Ministère de l'Intérieur, Paris. Disponible en https://www.interieur.gouv.fr/Archives/Archives-des-actualites/2016-Actualites/Lancement-de-l-application-mobile-SAIP (consultado el 14 de marzo de 2021).

[134] CADA. Opinión 20163619, de 20 de octubre de 2016.

[135] CADA. Opinión 20200496, de 12 de marzo de 2020.

[136] CTBG. R/0377/2017, de 30 e octubre, FFJJ 3 y 4.

[137] VEALE, VAN KLEEKS y BINNS (2018). "Fairness and Accountability…". *Op. cit*., pág. 7; DE LAAT, Paul B. (2018). "Algorithmic Decision-Making…". *Op. cit.*, pág. 539; BRAUNEIS, R. y GOODMAN, E. P. (2018). "Algorithmic Transparency for the Smart City". *The Yale Journal of Law & Technology*, vol. 20, pág. 160.

[138] Real Decreto 3/2010, de 8 de enero, por el que se regula el Esquema Nacional de Seguridad en el ámbito de la Administración Electrónica. El ENS es aplicable desde el punto de vista subjetivo a la Administración General del Estado, a las Administraciones de las Comunidades Autónomas y a las Entidades que integran la Administración Local, así como a las entidades de derecho público vinculadas o dependientes de las mismas, a las Universidades Públicas, a los ciudadanos en sus relaciones con las Administraciones Públicas, a las relaciones entre las distintas Administraciones Públicas (cfr. arts. 2, 3 y 156 de la Ley LRJSP y arts. 2 y 13 de la LPAC). Desde el punto de vista objetivo, el ENS se aplica a las Sedes y Registros electrónicos, los Sistemas de Información accesibles electrónicamente por los ciudadanos; Sistemas de Información para el ejercicio de derechos; Sistemas de Información para el cumplimiento de deberes; Sistemas de Información para recabar información y estado del procedimiento administrativo. Vid. CCN-CERT (s.f), *Esquema Nacional de Seguridad – Preguntas Frecuentes*. Disponible en https://www.ccn-cert.cni.es/publico/dmpublidocuments/ENS-FAQ.pdf(consultado el 14 de marzo de 2021).

[139] Art. 27, 43 y Anexo II del ENS.

[140] Vid. Art. 43.3 y Anexo II, apartado 5.8.2. del ENS con relación a la protección de servicios y aplicaciones web.

[141] GAIP. Resolución 200/2017, de 21 de junio, FJ 2. El solicitante en cuestión del código fuente era un profesor que había participado en diversas ocasiones y sin éxito en el procedimiento de selección de miembros de los Tribunales de Corrección de la PAU.

[142] CTBG. Resolución 701/2018, de 18 de febrero de 2019, FD 4.

[143] Vid. la Disposición Transitoria Segunda del Real Decreto 897/2017, de 6 de octubre, y la Resolución de 15 de noviembre de 2017, de la Secretaría de Estado de Energía, por la que se pone en marcha la aplicación telemática que permita al comercializador de referencia

comprobar que el solicitante del bono social cumple los requisitos para ser considerado consumidor vulnerable.

[144] El art. 8.2 del Real Decreto 919/2014, de 31 de octubre, por el que se aprueba el Estatuto del Consejo de Transparencia y Buen Gobierno, dispone que, entre las competencias del Presidente del CTBG, está *recabar de las distintas Administraciones Públicas la información necesaria para el cumplimiento de sus funciones*.

[145] Art. 2.1.a) del Real Decreto 421/2004, de 12 de marzo, por el que se regula el Centro Criptológico Nacional.

[146] CENTRO CRIPTOLÓGICO NACIONAL (s.f.). *Informe técnico sobre el incremento de riesgo asociado a la revelación del código fuente de las aplicaciones informáticas*. Disponible en https://civio.app.box.com/s/ufrg58o4z70nnm3j1q4j721oyur6je7a. Informe incorporado al escrito de 28 de enero de 2020 presentado por la Abogacía del Estado para su incorporación a los Autos en el Procedimiento Ordinario 18/2019.

[147] La Directiva (UE) 2016/1148 del Parlamento Europeo y del Consejo, de 6 de julio de 2016, relativa a las medidas destinadas a garantizar un elevado nivel común de seguridad de las redes y sistemas de información en la Unión (Directiva NIS) impone a los Estados Miembros la obligación de designar autoridades nacionales competentes, puntos de contacto únicos y CSIRT con funciones relacionadas con la seguridad de las redes y sistemas de información.

[148] CCN-CERT, 2018*. Vulnerabilidad en la librería libssh de código abierto*, de 25 de octubre. Disponible en https://www.ccn-cert.cni.es/seguridad-al-dia/avisos-ccn-cert/7133-ccn-cert-av-39-18-vulnerabilidad-en-la-libreria-libssh-de-codigo-abierto.html (consultado el 14 de marzo de 2021); INCIBE-CERT (2016), *FossHub hackeado e infectados la mayoría de sus programas*, de 2 de agosto. Disponible https://www.incibe-cert.es/alerta-temprana/bitacora-ciberseguridad/fosshub-hackeado-e-infectados-mayoria-sus-programas (consultado el 14 de marzo de 2021).

[149] Entre las vulnerabilidades más frecuentes de las aplicaciones web, destacarían, por ejemplo, la "inyección SQL" que consiste un *método de infiltración de código intruso que se sirve de una vulnerabilidad presente en una aplicación en el nivel de validación de entradas para la realización de consultas a una base de datos*; o las referencias directas a objetos inseguras (fichero, directorio, registro de la BD, clave…) en forma de URL o de parámetro de un formulario, con lo que un atacante podría manipular estas referencias para acceder a otros objetos sin autorización. Piénsese que entre las metodologías de análisis de vulnerabilidades de las aplicaciones web, el *análisis de caja blanca* implicaría precisamente el acceso completo al código fuente de la aplicación en busca de funciones vulnerables, incluso de documentación adicional relativa al lenguaje de programación de la aplicación Web, tal como funciones y librerías. Vid. CCN-CERT (2011). *Guía de Seguridad de las TIC. (CCN-STIC-812) Seguridad en entornos y aplicaciones web*, pág. 8 y 23; (2015). *Guía de Seguridad (CCN-STIC-401). Glosario y Abreviaturas*, pág. 560.

[150] CITRON (2008). "Technological Due Process". *Op. cit.*, pág. 1293; DIAKOPULOS, N. (2013). Algorithmic accountability reporting: on the investigation of black boxes, Tow Center for Digital Journalism, Columbia University, pág. 12; FINK (2018). "Opening the government's black boxes…". *Op. cit*., págs. 1454-1455.

[151] ETALAB (2021). *Ouvrir les codes sources*, 11 de febrero, pág. 8.

[152] CADA. Dictamen 20180226, de 17 de mayo de 2018.

[153] CADA. Opinión 20180376, de 31 de mayo de 2018. En este caso, el código fuente solicitado había sido desarrollado por investigadores asignados al OFCE sujetos a un régimen contractual de Derecho privado.

[154] Resolución del Parlamento Europeo 2020/2012(INL), apartado 86:

> *[…] cuando los fondos procedentes de fuentes públicas contribuyan significativamente al desarrollo, el despliegue o el uso de la inteligencia artificial, la robótica y las tecnologías conexas, junto con las normas abiertas de licitación y contratación, puede estudiarse la posibilidad de que sean públicos por defecto, previo acuerdo con el desarrollador, el código, los datos generados (en tanto en cuanto no sean personales) y el modelo formado, a fin de garantizar la transparencia, mejorar la ciberseguridad y posibilitar su reutilización, para fomentar la innovación* [cursiva nuestra].

[155] MANCOSU (2019). "Les algorithmes publics déterministes…" *Op. cit.*, págs. 77.

[156] GAIP. Resolución 200/2017, de 21 de junio, FJ 2: *Tampoco se ha invocado el límite relativo a los derechos de propiedad intelectual o industrial (artículo 21.1.g LTAIPBG), seguramente porque, como se ha señalado en el fundamento jurídico anterior, el programa informático en cuestión es de titularidad de la Generalitat.*

[157] GAIP, Resolución 200/2017, de 21 de junio, FJ 3. En sentido similar, véase la Resolución de 21 de septiembre de 2016, FJ 3:

> *Sin embargo, dada la finalidad de control de la petición de acceso, parece prudente restringir el acceso a este fin y no permitir la difusión o utilización del algoritmo sin la autorización expresa del Consejo Interuniversitario o de quien, eventualmente, ostente la titularidad del derecho de propiedad inmaterial en cuestión*.

[158] El acceso condicionado constituye una solución procedimental que, en virtud del principio de proporcionalidad, procedería aplicar cuando, prevaleciendo un interés público superior favorable al acceso, sin embargo, existen derechos e intereses legítimos de la propia Administración o de terceros que pueden verse afectados no tanto por el acceso, sino por una eventual difusión pública o explotación ilegítima de la información objeto del acceso por parte del solicitante. En tales casos, la materialización del acceso puede realizarse mediante consulta presencial de la información, pues ésta resulta menos lesiva del derecho a la propiedad intelectual de su autor al evitar riesgos de difusión no autorizada. En algunos casos, la consulta presencial puede someterse a determinadas cautelas como, por ejemplo, la advertencia expresa de la responsabilidad en que puede incurrir el solicitante con determinados usos personales de

la información entregada; la imposición de un deber de reserva o confidencialidad respecto de la información consultada, así como el establecimiento, en su caso, de medidas necesarias para evitar el uso de dispositivos móviles aptos para la obtención de copias. GAIP. Dictamen 1/2016, de 11 de mayo, FFJJ 2 y 5; Resolución 261/2017, de 26 de julio, FJ 2; Reclamación 62/2017, de 22 de febrero, FJ 3; Reclamación 17/2015, de 23 de diciembre, FJ 5.

[159] ICO. FS50630372, de 18 de julio de 2019, paras. 2, 3, 10-12. El contexto de la solicitud tiene que ver con dos aplicaciones, Cysill y Cysgeir, que forman parte del paquete de software, Cysgliad, cuyos derechos de propiedad intelectual, incluido el derecho *sui generis* sobre las bases de datos, corresponden a la Universidad de Bangor. Cysill incluye un revisor ortográfico y gramatical y tesauro. Cysgeir comprende diccionarios electrónicos en Galés/Inglés. Cysgliad es un paquete de software que combina Cysill y Cysgir con distintas funcionalidades de revisión ortográfica, gramatical y mutaciones lingüísticas, diccionarios, tesauros, y conjugador de verbos. El desarrollo original del software Cysgliad para PC fue parcialmente financiado por el Consejo de la Lengua Galesa (2004) y, posteriormente, se desarrolló la versión para Mac (2007). Sin embargo, desde entonces la Universidad no recibió más ayudas públicas para la versión PC del software, por lo que las actualizaciones necesarias corren de cuenta del personal de la Unidad de Tecnologías para la Lengua de la propia Universidad. Pues bien, el reclamante solicitó el acceso a los códigos fuente de las aplicaciones Cysill y Cysgeir para Mac, así como los códigos fuente de los diccionarios y bases de datos relacionadas. Respecto de las aplicaciones, la Universidad de Bangor contestó que sus respectivos códigos fuente estaban disponibles públicamente en la propia página web remitiendo al solicitante a la URL concreta. Sin embargo, la entidad reclamada desestimó el acceso a los códigos fuente para los diccionarios y bases de datos relacionadas pues constituían *información comercial sensible* según la sección 43(2) de la FOIA2000.

[160] *Idem*, paras. 43-44, 51.

[161] *Idem*. paras. 16, 38.

[162] *Idem*, paras. 10 y 23. Las versiones para Mac de los códigos fuente de Cysill y Cysgeir fueron distribuidas a través de una licencia *open source* "BSD" (*Berkeley Software Distribution*) desarrollada por la Universidad de Berkeley para sistemas operativos BSD basados en Unix.

[163] *Idem*, paras. 13-15, 28. Así, por ejemplo, se explica en la resolución que componentes clave del software Cysgliad y de sus bases de datos habían sido licenciados separadamente a Microsoft para su corrector ortográfico de galés y a la BBC para su diccionario *LearnWelsh* y juegos de aprendizaje de la lengua. Asimismo, todos los ingresos obtenidos a través de la venta de licencias de Cysgliad para PC se destinaban íntegramente al pago del sueldo del personal de la Unidad, que sólo recibe financiación interna de la Universidad.

[164] *Idem*, para. 47.

[165] *Idem*, paras. 18, 19, 28. Según queda acreditado en la Resolución, resulta significativo que, a través de distintos posts publicados en redes sociales por el solicitante, la Universidad fue conocedora de que éste había indicado su intención de desarrollar *languagetool*, un producto rival de Cysill. La Universidad explicó que languagetool (https://languagetool.org/) se refiere a un corrector ortográfico y gramatical de código abierto específico que es similar a su software Cysill que la comunidad de código abierto puede desarrollar para cualquier idioma. Languagetool necesita reglas y datos específicos del idioma para que funcione con un idioma

específico. La Universidad considera que, en realidad, este es el propósito que hay detrás de la solicitud. De hecho, el sitio web languagetool.org está operativo y ofrece servicios premium de pago.

[166] *Idem*, para. 52.

[167] Consta también la Resolución R/0275/2016, de 25 de agosto, que fue objeto de archivo por desistimiento tácito del reclamante de acceso al no haber subsanado en plazo un defecto formal de su reclamación ante el Consejo. La reclamación ante el CTBG trae causa de una solicitud que tuvo entrada el 11 de abril de 2016 en el Portal de Transparencia del Gobierno y que tenía objeto el acceso al código fuente de LexNET. Argumentaba el solicitante que, ante el elevado número de incidencias que había generado la aplicación, la finalidad de su petición era "comprobar con ayuda de un perito que [dicha aplicación] se encuentra bien desarrollada técnicamente y no genera problemas de seguridad". La respuesta desestimatoria de la Secretaría de Estado para la Administración de Justicia indicaba que: "[…] LexNET es una marca registrada por el Ministerio de Justicia en el Registro de Propiedad Industrial, por lo que dicha petición incurre en el límite del derecho de acceso regulado en el artículo 14.1.j) de la Ley 19/2013 de 9 de diciembre, que dispone que «el derecho de acceso podrá ser limitado cuando acceder a la información suponga un perjuicio para (…): j) El secreto profesional y la propiedad intelectual e industrial»".

[168] La jurisprudencia comunitaria ha aclarado que el objeto de la protección conferida por la Directiva 91/250/CEE del Consejo, de 14 de mayo de 1991 (actualmente, sustituida por la DIRECTIVA 2009/24/CE) *abarca el programa de ordenador en todas sus formas de expresión, que permiten reproducirlo en diferentes lenguajes informáticos, tales como el código fuente y el código objeto*; y, en general, todas *las formas de expresión de un programa de ordenador así como los trabajos preparatorios de concepción que pueden llevar respectivamente a la reproducción o a la creación ulterior de tal programa*. Sin embargo, tal protección no alcanzaría a la interfaz gráfica de usuario, que permite una comunicación entre el programa de ordenador y el usuario, en la medida en que esta interfaz *no permite reproducir ese programa de ordenador, sino que solo constituye un elemento de dicho programa por medio del cual los usuarios utilizan las funcionalidades de éste*. Véase, la STJUE (Sala Tercera) de 22 de diciembre de 2010, Asunto C-393/09, Bezpečnostní softwarová asociace, apartados 34-42. Asimismo, *ni la funcionalidad de un programa de ordenador ni el lenguaje de programación o el formato de los archivos de datos utilizados en un programa de ordenador para explotar algunas de sus funciones constituyen una forma de expresión de tal programa*. Véase, la STJUE (Gran Sala) de 2 de mayo de 2012, asunto C-406/10, SAS Institute, apartados 35-42. Dicho de otro modo, si el límite de la propiedad intelectual es procedente y aplicable al código fuente igual suerte deberían haber corrido las *especificaciones técnicas*. A menos que las especificaciones técnicas a las que se refiere el CTBG tengan un significado diferente al de la *documentación técnica* que protege también la legislación de propiedad intelectual, junto al código fuente.

[169] Cfr. DE LA CUEVA GONZÁLEZ-COTERA, J. (2019), "La configuración del software como cuestión política". *Teknokultura. Revista de Cultura Digital y Movimientos Sociales*, vol. 16, núm. 2, pág. 165. De opinión similar es Andrés Boix quien considera que

> *los algoritmos empleados por la Administración pública de modo no puramente instrumental producen materialmente los mismos efectos que cualquier*

> *reglamento, al preordenar la decisión final del poder público y limitar el ámbito de discreción o de capacidad de determinación de quienes los han de aplicar a partir de los postulados contenidos en la programación*.

Establecida esta identidad en sentido jurídico material entre algoritmos y reglamentos, los primeros han de ser tratados como tales normas reglamentarias *por el Derecho a la hora de regular cómo se producen, aplican y las garantías en torno a estos procesos*. BOIX PALOP, A. (2020). "Los algoritmos son reglamentos: la necesidad de extender las garantías propias de las normas reglamentarias a los programas empleados por la administración para la adopción de decisiones". *Revista de Derecho Público: Teoría y Método*, vol. 1., pág. 237.

[170] CONSEJO DE EUROPA (2018). *Algorithms and human rights. Study on the human rights dimensions of automated data processing techniques (in particular algorithms) and possible regulatory implications. Committee of Experts on Internet Intermediaries*, (MSI-NET). DGI(2017)12, pág. 37.

[171] AMMISTÍA INTERNACIONAL, ACCESS NOW (2018), *The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems* apartado 32; ACCESS NOW (2018). *Human Rights in the Age of Artificial Intelligence*, pág. 33.

[172] HOUSE OF LORDS (2018). *AI in the UK: ready, willing and able?* Select Committee on Artificial Intelligence, Report of Session 2017–19, apartado 95-99. MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL (2020). ENIA. *Estrategia Nacional de Inteligencia Artificial* (noviembre), págs. 57 y 58.

[173] DULONG DE ROSNAY, M. (2016). "Algorithmic Transparency and Platform Loyalty or Fairness in the French Digital Republic Bil". *Media Policy Project*, sp.; DATTA, A.; SEN, S. y ZICK, Y. (2016). "Algorithmic transparency via quantitative input influence: theory and experiments with learning systems". *IEEE Symposium on Security and Privacy*, pág. 598; BRAUNEIS, R. y GOODMAN, E. P. (2018). "Algorithmic Transparency…", *Op. Cit*., págs. 104, 110, 135. En el caso de la doctrina española, véanse, por ejemplo, las Conclusiones del I Seminario Internacional Derecho Administrativo e Inteligencia Artificial de 1 de abril de 2019 (Universidad de Castilla-La Mancha), en donde se afirma que: […] *se evidencia una llamativa falta de transparencia algorítmica y la ausencia de una adecuada percepción por las Administraciones Públicas sobre de la necesidad de aprobación de un marco jurídico específico* [subrayado nuestro]; ARELLANO TOLEDO, W. (2019). "El derecho a la transparencia algorítmica en Big Data e Inteligencia Artificial". *Revista General de Derecho Administrativo*, núm. 50, págs. 14-31 (aunque la autora plantea la transparencia algorítmica no tanto como un principio sino como un derecho); COTINO HUESO, L. (2019). "Riesgos e impactos del Big Data, la inteligencia artificial y la robótica. enfoques, modelos y principios de la respuesta del Derecho". *Revista General de Derecho Administrativo*, núm. 50, págs. 35-36; (BOIX (2020). "Los algoritmos son reglamentos…", Op. cit., pág. 260; LLANEZA, P. (2018), "Dataísmo, transparencia y protección de datos", RODRÍGUEZ MARÍN, S. y MUÑOZ GARCÍA, A. (Coords.) *Aspectos Legales de la Economía Colaborativa*, Madrid: Bosch, Wolters Kluwers, págs. 213-214.

[174] GÓMEZ JIMÉNEZ, M. L. (2019), *Urbanismo Participativo y Gobernanza Urbana en las Ciudades Inteligentes: el Efecto Reina Roja en Derecho Administrativo*. Navarra: Thomson-Reuters Aranzadi, págs. 179-180. La autora explica que el efecto "Reina Roja" pone de manifiesto cómo la acción se prioriza frente a la reflexión o al necesario debate ético que debe tener su

proyección en el debate jurídico, esperando respuestas inmediatas del Derecho ante los rápidos cambios sociales y disruptivos que propiciarían las tecnologías como el Big Data y los algoritmos de aprendizaje automatizado.

[175] Cfr. DIAKOPULOS, N. (2013). *Algorithmic accountability reporting: on the investigation of black boxes*, Tow Center for Digital Journalism, Columbia University, pág. 11; CERRILLO I MARTÍNEZ, A. (2019). El impacto de la inteligencia artificial en el Derecho Administrativo ¿nuevos conceptos para nuevas realidades técnicas?. *Revista General de Derecho Administrativo*, núm. 50, pág. 19.

[176] Cfr. CASTRO, D. (2018). *How Policymakers Can Foster Algorithmic Accountability*, Center for Data Innovation, pág. 8. Los autores identifican en el contexto norteamericano distintas aportaciones donde se defiende la apertura y acceso público a los modelos algorítmicos de IA.

[177] KOENE *et al.* (2019). "A governance framework…". *Op. cit*., pág. 30; NEW y CASTRO, D. (2018). *How Policymakers…*, *Op. cit*., pág. 5; DE LAAT (2017). "Algorithmic Decision-Making…" *Op. cit.*, págs. 527, 533-538; KROLL, J. A.; HUEY, J.; BAROCAS, S.; FELTEN, E. W., REIDENBERG, J. R.; ROBINSON, D. G.; YU, H. (2017). "Accountable Algorithms". *University of Pennsylvania Law Review*, vol. 165, núm. 3, pág. 658-660; EDWARDS, L. y VEALE, M. (2017). "Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for". *Duke Law & Technology Review*, vol. 16, núm. 18, pág. 8; ANANNY, M. y CRAWFORD, K. (2016). "Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability". *New Media & Society*, pág. 983; CERRILLO I MARTÍNEZ (2019). "El impacto de la inteligencia artificial..." *Op. cit*., pág. 18; PONCE SOLÉ, J. (2019). "Inteligencia Artificial, Derecho Administrativo y Reserva de Humanidad: Algoritmos y Procedimiento Administrativo Debido Tecnológico». *Revista General de Derecho Administrativo*, núm. 50, págs. 35-36.

[178] DEEKS, A. (2019). "The Judicial Demand for Explainable Artificial Intelligence. *Columbia Law Review*, vol. 119, núm. 7, pág. 1837.

[179] KROLL, J. A. *et al* (2017). "Accountable Algorithms"…, *Op. cit*., pág. 638.

[180] HOUSE OF LORDS (2018). *AI in the UK… Op. cit.,* apartado 96.

[181] ANANNY y CRAWFORD (2016). "Seeing without knowing…", *Op. cit.*, págs. 981-982.

[182] FINK, K. (2018), "Opening the government's black boxes". *Op. cit*., pág. 1454; PEREL, M. y ELKIN-KOREN, N. (2017). "Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement". *Florida Law Review*, vol. 69, núm. 181, págs. 184-185.

[183] Vid. CONSEJO DE EUROPA (2018). *Algorithms and human rights. Study on the human rights dimensions of automated data processing techniques (in particular algorithms) and possible regulatory implications. Committee of Experts on Internet Intermediaries*, (MSI-NET). DGI(2017)12, págs. 35-37; COMISIÓN EUROPEA (2017). *Tender Specifications. Study on Algorithmic Awareness Building SMART 2017/0055*, DG for Communications Networks, Content and Technology, 15 de julio, pág. 6. En ambos casos, tanto el Consejo de Europa como la Comisión utilizan la expresión de "transparencia efectiva".

[184] ANANNY, M. y CRAWFORD, K. (2016). "Seeing without knowing…", *Op. cit.*, pág. 982.

[185] BOIX (2020). *Los algoritmos son reglamentos*. *Op. cit*., pág. 242.

[186] VALERO TORRIJOS, J. (2019). "Las garantías jurídicas de la inteligencia artificial en la actividad administrativa desde la perspectiva de la buena administración». *Revista Catalana de Dret Públic*, vol. 58, pág. 89.

[187] Ley 30/1992, de 26 de noviembre, de Régimen Jurídico de las Administraciones Públicas y del Procedimiento Administrativo Común.

[188] PONCE SOLÉ, J. (2019). "Inteligencia artificial, Derecho administrativo…", *Op. cit.*, pág. 17.

[189] VALERO TORRIJOS, J. (2018). "La tramitación del procedimiento administrativo por medios electrónicos". En ALMEIDA, M.; MÍGUEZ, L. (Dirs.) *La actualización de la administración electrónica*, Santiago de Compostela: Andavira, pág. 236; CERRILLO I MARTÍNEZ, A. (2019). «Com obrir les caixes negres de les administracions públiques? Transparència i rendició de comptes en l'ús dels algoritmes». *Revista Catalana de Dret Públic*, núm. 58, págs. 19-20.

[190] COGLIANESE y LEHR, D. (2017). "Regulating by Robot…", *Op. cit*., págs. 1206-1207.

[191] BARREDO, A.; DÍAZ-RODRÍGUEZ, N.; DEL SER, J., *et al.* (2020). "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI". *Information Fusion*, vol. 58, pág. 100; CÁTEDRA iDANAE (2019). *Interpretabilidad de los Modelos de Inteligencia Artificial, Universidad Politécnica de Madrid*, Management Solutions, Newsletter Trimestral, 3T2019, pág. 4.

[192] GUNING, D. (2017). *Explainable Artificial Intelligence* (XAI), [7] D. Technical Report, Defense Advanced Research Projects Agency (DARPA).

[193] BARREDO *et al*. (2020), "Explainable Artificial Intelligence (XAI)…". *Op. cit*., pág. 84; CARVALHO, D. V.; PEREIRA, E. M.; CARDOSO, J. S. (2019). "Machine Learning Interpretability: A Survey on Methods and Metrics". *Electronics*, vol. 8, núm. 8: 832, pág. 10; MOLNAR, C. (2018), *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. Leanpub, sp.

[194] CÁTEDRA iDANAE (2019). *Interpretabilidad*…, Op. cit., 3.

[195] BARREDO *et al*. (2020), "Explainable Artificial Intelligence (XAI)…". *Op. cit*., pág. 84.

[196] CARVALHO *et al.* (2019). "Machine Learning Interpretability…" *Op. cit*., pág. 15.

[197] MITTELSTADT, B.; RUSSELL, C.; WACHTER, S. (2019). "Explaining Explanations in AI". *FAT*'19: Proceedings of the Conference on Fairness, Accountability, and Transparency*, pág. 280; DEEKS (2019). "The judicial demand…" págs. 1832; BARREDO *et al.* (2020). "Explainable Artificial Intelligence (XAI)…", *Op. cit*., pág. 83.

[198] LIPTON, Z. C. (2018). "The Mythos of Model Interpretability". *ACM Queue*, vol. 16, núm. 3 (mayo-junio), pág. 12; LEPRI, B.; OLIVER, N. *et al*. (2018). "Fair, transparent and accountable algorithmic decision-making processes. The premise, the proposed solutions, and the open challenges". *Philosophy & Technology*, vol. 31, núm., 4, pág. 9; BARREDO *et al*. (2020).

Explainable Artificial Intelligence (XAI)… Op. cit., págs. 88-100; ICO y ALAN TURING INSTITUTE (2020), *Explaining decisions… Op. cit*. págs. 61-63, 115-118.

[199] MITTELSTADT, B.; RUSSELL, C.; WACHTER, S. (2019). "Explaining Explanations…", *Op. cit.*, pág. 280.

[200] BARREDO *et al*. (2020). *Explainable Artificial Intelligence (XAI)… Op. cit*, pág. 90.; ICO y ALAN TURING INSTITUTE (2020). *Explaining decisions…*, *Op. cit*., págs. 67-68.

[201] LEPRI, B.; OLIVER, N. et al. (2018). "Fair, transparent…", *Op. cit*., pág. 12; CARVALHO (2019). "Machine Learning Interpretability…" *Op. cit*., pág. 12; MOLNAR (2018). *Interpretable Machine Learning*, *Op. cit.*, sp.

[202] ICO y ALAN TURING INSTITUTE (2020). *Explaining decisions…*, *Op. cit*., pág. 69.

[203] MITTELSTADT *et al*. (2019). "Explaining Explanations…", Op. cit., pág. 280; LIPTON (2018). "The Mythos…", *Op. cit*, págs. 15-19; BARREDO *et al*. (2020). "Explainable Artificial Intelligence (XAI)…", *Op. cit*., pág. 88; ICO y ALAN TURING INSTITUTE (2020). *Explaining decisions… Op. cit*., pág. 120; MOLNAR (2018). *Interpretable Machine Learning*, *Op. cit.*, sp.

[204] CARVALHO (2019). "Machine Learning Interpretability…" *Op. cit.*, pág. 12-13.

[205] ICO y ALAN TURING INSTITUTE (2020). *Explaining decisions… Op. cit*., pág. 69; DEEK (2019). "The judicial demand…", Op. cit., pág. 1836; CARVALHO (2019). "Machine Learning Interpretability…" *Op. cit*., págs. 14-15.

[206] Nótese, por ejemplo, que la Resolución del Parlamento Europeo, de 20 de octubre de 2020 ((2020/2012(INL)), en su apartado 20, advierte de que debe tenerse en cuenta *la importante distinción entre la transparencia de los algoritmos y la transparencia en el uso de los algoritmos* (subrayado, nuestro).

[207] Vid. Exposición de motivos de la LTAIBG.

[208] Sentencia del Tribunal de Primera Instancia (Sala Cuarta), Kuijer/Consejo (T-211/00), de 7 de febrero de 2002, párrafo 52.

[209] PARLAMENTO EUROPEO (2017), (2015/2103(INL), *Op. cit.*, (*principios éticos*), núm. 12.

[210] Cfr. DE LAAT (2018). "Algorithmic Decision-Making…". *Op. cit.*, pág. 527.

[211] CERRILLO I MARTÍNEZ (2019). "Com obrir les caixes negres…", *Op. cit*., págs. 21-22.

[212] DESAI y KROLL (2017). "Trust but verify…". *Op. cit.*, págs. 10-11.

[213] KOENE *et al.* (2019). "A governance framework…". *Op. cit*., pág. 8.

[214] Véanse, entre otros los arts. 5.4 y 5.5 LTAIBG, arts. 5.5 y 6.1.c) de la Ley 19/2014, de 29 de diciembre, de Transparencia, Acceso a la Información Pública y Buen Gobierno de Cataluña; arts.

6.3, 11.1 y 11.3 de la Ley 8/2015, de 25 de marzo, de Transparencia de la Actividad Pública y Participación Ciudadana de Aragón; art. 6.5 de la Ley 1/2016, de 18 de enero, de Transparencia y Buen Gobierno de Galicia; art. 5.c) y art. 11.5 de la Ley Foral 5/2018, de 17 de mayo, de Transparencia, Acceso a la Información Pública y Buen Gobierno de Navarra; arts. 6.d), 7, 8.1.a) y 8.2 de la Ley 10/2019, de 10 de abril, de Transparencia y de Participación de la Comunidad de Madrid. Entre la legislación autonómica debe destacarse el art. 5.1 de la Ley 8/2018, de 14 de septiembre, de Transparencia, Buen Gobierno y Grupos de Interés, donde las exigencias de comprensibilidad sí que están conectadas a la contextualización de la información. En concreto el precepto dispone que la información sujeta a obligaciones de publicidad activa:

> *Será actualizada, veraz, coherente, estructurada, concisa, completa, segura, de fácil acceso, multicanal, comparable, multiformato, interoperable, reutilizable, entendible y clara con resúmenes, textos introductorios, glosarios terminológicos, fichas, cuadros sinópticos y elementos análogos que ayuden a la comprensión de la información por el ciudadano medio* [subrayado, nuestro].

[215] MESEGUER YEBRA, J. *et al*. (2017). *Comentarios sobre aspectos clave en materia de acceso a información pública*. Federación Española de Municipios y Provincias y Red de Entidades Locales por la Transparencia y por la Participación Ciudadana. Navarra: Thomson Reuters Aranzadi, págs. 95-96.

[216] CERRILLO I MARTÍNEZ (2019). "Com obrir les caixes negres…", *Op. cit*., págs. 21-22.

[217] En la mayoría de las recomendaciones y propuestas analizadas las obligaciones de transparencia y suministro de información parecen identificarse exclusivamente con relación a las Autoridades públicas de control o supervisión que, en su caso, se establezcan y, de forma muy limitada tales obligaciones se formulan respecto de los interesados o de la ciudadanía en general.

Véanse los principios para la administración responsable de una IA confiable (apartado 1.3):

> *Los actores de IA deben comprometerse con la transparencia y la difusión responsable respecto a los sistemas de IA. Con este fin, deben proporcionar información significativa, adecuada al contexto y coherente con el estado de la técnica para: (i) fomentar una comprensión general de los sistemas de IA; (ii) concienciar a las partes interesadas de sus interacciones con los sistemas de IA, incluso en el lugar de trabajo; (iii) Permitir que los afectados por un sistema de IA comprendan el resultado, y (iv). posibilitar que aquellos afectados negativamente por un sistema de inteligencia artificial cuestionen su resultado basándose en información fácil de entender sobre los factores y la lógica que sirvió de base para la predicción, recomendación o decisión*.

[219] La Recomendación dispone que:

> *Los Estados deben establecer niveles apropiados de transparencia con respecto a la contratación pública, el uso, el diseño y los criterios y métodos básicos de procesamiento de los sistemas algorítmicos implementados por y para ellos, o por actores del sector privado. Los marcos legislativos para la propiedad intelectual o los secretos comerciales no deben impedir tal transparencia, ni los Estados o las partes privadas deben tratar de explotarlos para este fin. Los niveles de transparencia deben ser lo más altos posible y proporcionales a la gravedad de los impactos adversos sobre los derechos humanos, incluidas etiquetas o sellos éticos para los sistemas algorítmicos que permitan a los usuarios navegar entre sistemas. El uso de sistemas algorítmicos en los procesos de toma de decisiones que conllevan un alto riesgo para los derechos humanos debe estar sujeto a estándares particularmente altos en lo que respecta a la explicabilidad de los procesos y resultados.*

[220] La Resolución del Parlamento incluye una Propuesta de Reglamento sobre los principios éticos para el desarrollo, el despliegue y el uso de la inteligencia artificial, la robótica y las tecnologías conexas, en suyo considerando (21) se afirma:

> *Para garantizar la transparencia y la rendición de cuentas, se debe informar a los ciudadanos siempre que un sistema utilice inteligencia artificial, siempre que los sistemas de inteligencia artificial personalicen un producto o un servicio para sus usuarios, así como de si pueden desactivar o limitar la personalización, y siempre que se enfrenten a una tecnología de toma de decisiones automatizada. Además, las medidas de transparencia deben ir acompañadas, siempre que sea técnicamente posible, de explicaciones claras y comprensibles sobre los datos utilizados y el algoritmo, así como sobre su finalidad, sus resultados y sus riesgos potenciales.*

Sin embargo, la Propuesta no determina cómo se concretaría ese deber de información a los ciudadanos.

[221] Aquí el principio de transparencia se identifica con la trazabilidad (obligación de documentar *con arreglo a la norma más rigurosa posible* el origen y etiquetado de los datos, los procesos, los algoritmos utilizados y las decisiones adoptadas por el modelo), la explicabilidad (capacidad de explicar tanto los procesos técnicos de un sistema de IA como las decisiones humanas asociadas mediante información oportuna según el nivel de especialización de la parte interesada) y la comunicación (derecho de las personas a saber si están interactuando con un sistema de IA y a decidir si quiere interactuar, en su caso, con una persona).

[222] Por ejemplo, en el art. XVI.7c) de la Carta se reconoce el derecho específico de los interesados con relación a la IA en sus relaciones con la Administraciones públicas a *obtener una motivación comprensible en lenguaje natural de las decisiones que se adopten en el entorno digital, con*

justificación de las normas jurídicas relevantes al caso y de los criterios de aplicación de las mismas. Y en el art. XXIII.1b) se dispone que, en el desarrollo y ciclo de vida de los sistemas de IA –incluidos los implementados por las Administraciones públicas– deberán asegurarse *la transparencia, auditabilidad, explicabilidad y trazabilidad*.

[223] Cfr. COM(2020) 65 final, págs. 12, 17, 19, 23-24.

[224] Los sistemas de IA alto-riesgo vienen identificados mediante un sistema de *numerus clausus* previsto en el art. 6 con relación al listado del Anexo II (componentes de seguridad de los productos contemplados en la legislación sectorial de la Unión) y los del Anexo III (sistemas biométricos remotos de identificación y categorización de personas naturales en tiempo real o *ex post*; gestión y funcionamiento de infraestructuras críticas relacionadas con la gestión del tráfico y provisión de agua, gas, calor y electricidad; sistemas que determinen el acceso de las personas a instituciones de educación y formación vocacional; sistemas de selección y evaluación de candidatos en el ámbito del empleo y gestión de trabajadores; sistemas utilizados por las autoridades públicas competentes para la detección de *deep fakes*, la detección, prevención o persecución de delitos, elaboración de perfiles de riesgo criminal, evaluación de la fiabilidad de pruebas en el ámbito penal, migración, asilo y control de fronteras; sistemas de apoyo a las autoridades judiciales para la valoración de hechos y aplicación de la ley y procesos democráticos).

[225] La propuesta de Reglamento establece una serie de obligaciones relativas a la generación, conservación, registro, puesta a disposición de información relevante que garantizarían el principio de transparencia. En particular, estas obligaciones se concretan en la generación y conservación de la documentación técnica que acredite el cumplimiento de los requisitos impuestos en el Capítulo 2 del futuro Reglamento a los sistemas de alto-riesgo frente a las autoridades de control competentes (art. 11), el registro de eventos o *logs* del sistema (art. 12) y el suministro de información específica para la interpretación de los resultados del sistema y uso adecuado del mismo.

[226] Entre estas obligaciones de transparencia y suministro de información previstas en el art. 13.3 de la propuesta de Reglamento se encuentran: (a) la identidad y los datos de contacto del proveedor; (b) las características, capacidades y limitaciones de rendimiento del sistema, incluyendo: (i) su finalidad prevista; ii) el nivel de precisión, robustez y ciberseguridad; (iii) cualquier circunstancia conocida o previsible, relacionada con el uso del sistema que pueda conducir a riesgos para la salud y seguridad o los derechos fundamentales; (iv) su rendimiento en lo que respecta a las personas o grupos de personas en las que se pretenda utilizar el sistema; (v) cuando sea apropiado, las especificaciones relativas a datos de entrada, o cualquier otra información relevante relativa a los datos de entrenamiento, validación y prueba utilizados; los cambios en el sistema de IA de alto riesgo y su desempeño que hayan sido predeterminados por el proveedor en el momento de la evaluación de la conformidad inicial, si los hubiere; (d) las medidas de supervisión humana, incluidas las medidas técnicas establecidas para facilitar la interpretación de los resultados de los sistemas de inteligencia artificial por parte de los usuarios; (e) la vida útil prevista del sistema y las medidas de mantenimiento necesarias para garantizar el correcto funcionamiento del sistema de IA, incluidas las actualizaciones de software.

[227] Cfr. art. 13.1 y 2 de la propuesta de Reglamento, cuya versión única en inglés, dispone:

> *High-risk AI systems shall be designed and developed in such a way to ensure that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately. 2. High-risk AI systems shall be accompanied by instructions for use in an appropriate digital format or otherwise that include concise, complete, correct and clear information that is relevant, accessible and comprehensible to users* [subrayado, nuestro].

En los preceptos transcritos parece que se está pensando en la contratación externa del diseño y desarrollo de proyectos de IA por parte del cliente-usuario del sistema, lo cual, por otra parte, es común en el ámbito de las Administraciones y sector público, donde se recurre habitualmente a la licitación de estos proyectos en lugar de desarrollos *in-house*.

[228] DIE BUNDESBEAUFTRAGTE FÜR DEN DATENSCHUTZ UND DIE INFORMATIONSFREIHEIT (2018). *Transparenz der Verwaltung beim Einsatz von Algorithmen für gelebten Grundrechtsschutz unabdingbar*. 36 Konferenz der Informationsfreiheitsbeauftragten in Deutschland, pág. 4.

[229] DATENETHIKKOMMISSION (2019). *Gutachten der Datenethikkommission*, págs. 187 y 216.

[230] Decreto Nº 2017-330, de 14 de marzo de 2017, relativo a los derechos de las personas que sean objeto de decisiones individuales adoptadas sobre el fundamento de un tratamiento algorítmico.

[231] Sin perjuicio de la debida protección al deber de secreto previsto en la Ley en aquellos casos en que concurra un límite legal tasado (e.g. derechos de propiedad intelectual de terceros, el secreto deliberatorio, etc).

[232] DULONG DE ROSNAY (2016). "Algorithmic Transparency…", *Op. cit*., sp.

[233] HUERGO LORA, A. (2020). "Una aproximación a los algoritmos desde el Derecho Administrativo". HUERGO LORA, A. (2020). *La regulación de los algoritmos*, Navarra, Thomson-Reuters, Aranzadi, págs. 72-73. A nuestro juicio, sin embargo, la distinción propuesta por el autor puede ser harto complicada en la práctica, dada la creciente automatización y digitalización de los procesos. Es más, en muchos de los procedimientos judiciales que, en el Derecho comparado, están teniendo lugar, un lugar común en los mismos es que queda sin determinar si la decisión algorítmica discutida anulada se basa o no en un modelo determinista o en un modelo de IA. Como hemos visto en el caso de K.W. v. Armstrong (2016), comentado *supra*, en ningún momento, la resolución del Tribunal de Distrito de Idaho explicita si la aplicación informática que determinaba la cuantía de la ayuda pública para personas vulnerables se limitaba a mecanizar el procedimiento de concesión de la ayuda, o si entraba en juego algún tipo concreto de algoritmo de aprendizaje automatizado. Tampoco, en los asuntos resueltos por los Tribunales holandeses.

[234] Como señala además la *Memoria de Impacto Normativo* del Proyecto de Ley para una República Digital, este derecho de acceso a la información relativa a los tratamientos algorítmicos completa y refuerza el marco jurídico aplicable a las personas físicas en materia de

protección de datos. Cfr. LEGIFRANCE (2015). *Projet de Loi pour une République numérique. Etude d'Impact*. 9 de diciembre, págs. 10-12. En efecto, los arts. 13.2.f), 14.2.g), 15.1.h) y 22.3 RGPD exigen al responsable de tratamiento –lo que incluye también a las organizaciones públicas– comunicar al interesado la existencia de decisiones automatizas, incluida la elaboración de perfiles, la información significativa sobre la lógica aplicada, así como la importancia y las consecuencias previstas de dicho tratamiento para el interesado, al tiempo que reconoce el derecho del interesado a obtener intervención humana por parte del responsable, a expresar su punto de vista y a impugnar la decisión basada en tratamientos automatizados, incluida la elaboración de perfiles.

[235] ACCESS NOW (2018). *Human rights in the Age of Artificial Intelligence.*

[236] DE LA QUADRA-SALCEDO, T. (2018). "Retos, riesgos y oportunidades de la sociedad digital". En: DE LA QUADRA-SALCEDO y PIÑAR MAÑAS, *Op. cit*., RED.ES, pág.54

[237] PONCE SOLÉ, J.(2019). "Inteligencia artificial, Derecho administrativo y Reserva de Humanidad: Algoritmos y Procedimiento Administrativo Debido Tecnológico». *Revista General de Derecho Administrativo*, núm. 50, pág. 45.

[238] VALERO TORRIJOS (2018). "La tramitación del procedimiento administrativo…". *Op. cit*., pág. 237.

[239] KOENE *et al* (2019). "A governance framework…". *Op. cit.*, pág. 6;

[240] Cfr. *The Toronto Declaration. Protecting the right to equality in machine learning* (2018), apartado 32; KOENE *et al* (2019). "A governance framework…". *Op. cit.*, págs. 5-6; ICO y ALAN TURING INSTITUTE (2020). *Explaining decisions… Op. cit.*, 20, 26, 69, 120-122; ICO (2020). *Guidance on AI and Data Protection*, 30 July, v. 0.0.41.

Referencias

ACCESS NOW. (2018). *Human rights in the age of Artificial Intelligence.* Disponible en https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf (consultado el 14 de marzo de 2021)

AI NOW INSTITUTE (2018). *Litigating algorithms: challenging government use of algorithmic decision systems*, New York University. Disponible en https://ainowinstitute.org/litigatingalgorithms.pdf (consultado el 14 de marzo de 2021)

—*Taking algorithms to court. Current strategies for litigating government use of algorithmic decision-making*. Disponible en https://ainowinstitute.org/announcements/litigating-algorithms.html (consultado el 14 de marzo de 2021)

ALAMILLO DOMINGO, I. Y URIOS APARISI, X. (2011). *La actuación administrativa automatizada en el ámbito de las Administraciones Públicas. Análisis jurídico y metodológico para la construcción y la explotación de trámites automáticos*. Barcelona: Escola d'Administració Pública de Catalunya.

ANANNY, M. y CRAWFORD, K. (2016). «Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability». *New Media & Society*. DOI: 10.1177/1461444816676645 (consultado el 14 de marzo de 2021).

ARELLANO TOLEDO, W. (2019). «El derecho a la transparencia algorítmica en Big Data e Inteligencia Artificial». *Revista General de Derecho Administrativo*, núm. 50.

AUTORITAT CATALANA DE PROTECCIÓ DE DADES (APDCAT). (2020). *Intel·ligència Artificial. Decisions automatitzades a Catalunya*, Barcelona (enero). https://apdcat.gencat.cat/web/.content/04-actualitat/noticies/documents/INFORME-INTELLIGENCIA-ARTIFICIAL-FINAL-WEB-OK.pdf (consultado el 14 de marzo de 2021)

AUBY, J.-B. (2018). «Algorithmes et Smart Cities: Données Juridiques», *Revue Générale du Droit*, numéro 29878.

BABUTA, A.; OSWALD, M.; RINIK, C. (2018). *Machine Learning Algorithms and Police Decision-Making. Legal, Ethical and Regulatory Challenges*. RUSI Whitehall Report 3-18, Universidad de Winchester.

BARREDO, A.; DÍAZ-RODRÍGUEZ, N.; DEL SER, J.; BENNETOT, A.; TABIK, S.; BARBADO, A.; GARCÍA, S.; GIL-LÓPEZ, S.; MOLINA, D.; BENJAMINS, R.; CHATILA, R.; HERRERA, F. (2020). "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI". *Information Fusion*, vol. 58, pág. 83. https://doi.org/10.1016/j.inffus.2019.12.012 (consultado el 14 de marzo de 2021)

BASOGAIN, X. (s.f). *Redes neuronales artificiales y sus aplicaciones*. Escuela Superior de Ingeniería de Bilbao. Disponible en https://ocw.ehu.eus/pluginfile.php/40137/mod_resource/content/1/redes_neuro/contenidos/pdf/libro-del-curso.pdf (consultado el 14 de marzo de 2021)

BINNS, R. y GALLO, V. (2019). *Automated decision making: the role of meaningful human reviews*, Information Commissioner's Office.

BOIX PALOP, A. (2020). «Los algoritmos son reglamentos: la necesidad de extender las garantías propias de las normas reglamentarias a los programas empleados por la administración para la adopción de decisiones». *Revista de Derecho Público: Teoría y Método*, vol. 1. DOI: 10.37417/RPD/vol_1_2020_33 (consultado el 14 de marzo de 2021).

BORNSTEIN, S. (2018). "Antidiscriminatory algorithms". *Alabama Law Review*. 2018, Vol. 70, Issue 2.

BOURCIER, D. y DE FILIPPI, P. (2018). « Les algorithmes sont-ils devenus le langage ordinaire de l'administration? ». En: KOUBI, G. ; CLUZEL-METAYER, L. ; TAMZINI, W.. *Lectures critiques du Code des relations Public et Administration*, LGDJ, hal-01850928f.

– (2018). «La transparence des algorithmes face à l'Open Data: Quel statut pour les données d'apprentissage?» En: *Revue Française d'Administration Publique*, ENA, ffhal-01850926f.

CAPDEFERRO, Ó. (2019). «Las herramientas inteligentes anticorrupción: entre la aventura tecnológica y el orden jurídico». En *Revista General de Derecho Administrativo*, Nº 50.

CARVALHO, D. V.; PEREIRA, E. M.; CARDOSO, J. S. (2019). «Machine learning interpretability: A survey on methods and metrics». *Electronics*, vol. 8, núm. 8: 832. https://doi.org/10.3390/electronics8080832 (consultado el 14 de marzo de 2021)

CÁTEDRA IDANAE (2019). *Interpretabilidad de los modelos de Inteligencia Artificial,* Universidad Politécnica de Madrid, Management Solutions, Newsletter Trimestral, 3T2019.

CESEDEN (2013). *Big data en los entornos de Defensa y Seguridad*. Documento de investigación 03/2013. Instituto Español de Estudios Estratégicos.

CERRILLO I MARTÍNEZ, A. (2019). «Com obrir les caixes negres de les administracions públiques? Transparència i rendició de comptes en l'ús dels algoritmes». *Revista Catalana de Dret Públic*, núm. 58. Disponible en http://revistes.eapc.gencat.cat/index.php/rcdp/article/view/10.2436-rcdp.i58.2019.3277 (consultado el 14 de marzo de 2021)

— (2020) ¿Son fiables las decisiones de las Administraciones Públicas adoptadas por algoritmos?, *European Review of Digital Administration & Law*, Vol. 1, Issue 1-2, Erdal. DOI 10.4399/97888255389603.

— (2019). «El impacto de la inteligencia artificial en el Derecho Administrativo ¿nuevos conceptos para nuevas realidades técnicas?» *Revista General de Derecho Administrativo*, núm. 50.

CITRON, D.; CALO, R. (2019). *The Automated Administrative State*. Harvard Kennedy School, Shorenstein Center. Disponible en https://ai.shorensteincenter.org/ideas/2019/4/3/the-automated-administrative-state (consultado el 14 de marzo de 2021)

COGLIANESE, C. y LEHR, D. (2017). «Regulating by robot: Administrative decision making in the machine-learning era». *Georgetown Law Journal*, vol. 105, núm. 5. Disponible en https://scholarship.law.upenn.edu/faculty_scholarship/1734 (consultado el 14 de marzo de 2021)

COMMISSION D'ACCES AUX DOCUMENTS ADMINISTRATIF (2018). *Rapport d'activité 2018*. Disponible en https://www.cada.fr/sites/default/files/rapport_2018.pdf (consultado el 14 de marzo de 2021)

CONSEJO DE EUROPA (2018). *Algorithms and human rights. Study on the human rights dimensions of automated data processing techniques (in particular algorithms) and possible regulatory implications.* Committee of Experts on Internet Intermediaries, (MSI-NET). DGI(2017)12.

COTINO HUESO, L. (2020). «SyRI, ¿a quién sanciono? Garantías frente al uso de inteligencia artificial y decisiones automatizadas en el sector público y la sentencia holandesa de febrero de 2020». *LA LEY privacidad*, núm. 4, Wolters Kluwer, LA LEY 4999/2020.

—— (2019). «Riesgos e impactos del Big Data, la inteligencia artificial y la robótica. enfoques, modelos y principios de la respuesta del Derecho». *Revista General de Derecho Administrativo*, núm. 50.

DATENETHIKKOMMISSION (2019). *Gutachten der Datenethikkommission*. Disponible en https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf?__blob=publicationFile&v=6 (consultado el 14 de marzo de 2021)

DEEKS, A. (2019). «The judicial demand for explainable artificial intelligence. *Columbia Law Review*, vol. 119, núm. 7. Disponible en https://www.jstor.org/stable/26810851 (consultado el 14 de marzo de 2021)

DE LAAT, P. B. (2018). «Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?». *Philosophy & Technology*, núm. 31. Disponible en https://doi.org/10.1007/s13347-017-0293-z (consultado el 14 de marzo de 2021)

DE LA CUEVA, J. (2019). «La configuración del software como cuestión política». *Teknokultura. Revista de Cultura Digital y Movimientos Sociales*, vol. 16, núm. 2.

DE LA QUADRA-SALCEDO, T. (2018). «Retos, riesgos y oportunidades de la sociedad digital». DE LA QUADRA-SALCEDO, T.; PIÑAR MAÑAS, J. L. (Dirs.) *Sociedad Digital y Derecho*. Madrid: Boletín Oficial del Estado, Ministerio de Industria, Comercio y Turismo y RED.ES

DEPARTMENT FOR BUSINESS, ENERGY & INDUSTRIAL STRATEGY; DEPARTMENT FOR DIGITAL, CULTURE, MEDIA & SPORT; OFFICE FOR ARTIFICIAL INTELLIGENCE. (2020). *Guidelines for AI procurement*, sp. Disponible en https://www.gov.uk/government/publications/guidelines-for-ai-procurement/guidelines-for-ai-procurement (consultado el 14 de marzo de 2021)

DIAKOPOULOS, N. (2016). «Accountability in algorithmic decision making». *Communications of the ACM*, vol. 59, núm. 2, págs. 57-58. DOI:10.1145/2844110.

DIE BUNDESBEAUFTRAGTE FÜR DEN DATENSCHUTZ UND DIE INFORMATIONSFREIHEIT (2018). *Transparenz der Verwaltung beim Einsatz von Algorithmen für gelebten Grundrechtsschutz unabdingba*r. 36 Konferenz der Informationsfreiheitsbeauftragten in Deutschland. Disponible en https://www.datenschutz.rlp.de/fileadmin/lfdi/Konferenzdokumente/Informationsfreiheit/IFK/Entschliessungen/036_Algorithmen.pdf (consultado el 14 de marzo de 2021)

DOMINGOS, P. (2018). *The Master Algorithm. How the quest for the ultimate learning machine will remake our world*, NewYork: Basic Books.

DULONG DE ROSNAY, M. (2016). «Algorithmic Transparency and Platform Loyalty or Fairness in the French Digital Republic Bill». *Media Policy Project*. Disponible en http://eprints.lse.ac.uk/81295/ (consultado el 14 de marzo de 2021)

EDWARDS, L. y VEALE, M. (2017). «Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for». *Duke Law & Technology Review*, vol. 16, núm. 18.

ESTEVE PARDO, M. A. (2017). "El secreto profesional y la propiedad intelectual e industrial". En ARAGUAS GALERA, I. *et. al*. *Los límites al derecho de acceso a la información pública*. Madrid: INAP.

FJELD, J.; ACHTEN, N. *et al*. (2020). *Principled Artificial Intelligence: mapping consensus in ethical and rights-based approaches to principles for AI*. Cambridge (MA): Berkman Klein Center for Internet & Society at Harvard University.

GARCÍA DE ENTERRÍA, E. (2009). *Democracia, jueces y control de la Administración*, Navarra: Civitas Thomson Reuters.

GÓMEZ JIMÉNEZ, M. L. (2019). *Urbanismo participativo y gobernanza urbana en las ciudades inteligentes: el efecto Reina Roja en Derecho Administrativo*. Navarra: Thomson-Reuters Aranzadi.

GRUPO INDEPENDIENTE DE EXPERTOS DE ALTO NIVEL SOBRE INTELIGENCIA ARTIFICIAL (2019). *Directrices éticas para una IA fiable.* Bruselas: Comisión Europea.

GUNING, D. (2017). *Explainable Artificial Intelligence* (XAI), Technical Report, Defence Advanced Research Projects Agency (DARPA).

HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE (AI HLEG). *Ethics guidelines for trustworthy AI*. Brussels: European Commission, 8 de abril, pág. 21. Disponible en https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (consultado el 14 de marzo de 2021)

HUERGO LORA, A. (2020). *La regulación de los algoritmos*. Navarra: Thomson-Reuters, Aranzadi.

INFORMATION COMMISSIONER OFFICE (ICO) (2017). *Big data, artificial intelligence, machine learning and data protection*. Versión 2.2, de 4 de septiembre. Disponible en: https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf (consultado el 14 de marzo de 2021)

— (2020). *Explaining decisions made with AI*. Project explAIn (20 de Mayo), 1.0.312. Disponible en https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-artificial-intelligence-1-0.pdf (consultado el 14 de marzo de 2021)

— (2020). *Guidance on AI and data protection* (30 de julio), v. 0.0.41. Disponible en: https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/guidance-on-ai-and-data-protection/ (consultado el 14 de marzo de 2021)

KOENE, A.; CLIFTON, C.; HATADA, Y. *et al*. (2019). «A governance framework for algorithmic accountability and transparency», European Parliamentary Research Service, Scientific Foresight Unit (STOA). Disponible en https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf (consultado el 14 de marzo de 2021)

KROLL, J. A.; HUEY, J.; BAROCAS, S.; FELTEN, E. W., REIDENBERG, J. R.; ROBINSON, D. G.; YU, H. (2017). «Accountable algorithms». *University of Pennsylvania Law Review*, vol. 165, núm. 3.

LLANEZA, P. (2018), «Dataísmo, transparencia y protección de datos», RODRÍGUEZ MARÍN, S. y MUÑOZ GARCÍA, A. (Coords.) *Aspectos Legales de la Economía Colaborativa*, Madrid: Bosch, Wolters Kluwers, págs.. 213-214.

LEPRI, B.; OLIVER, N. *et al*. (2018). "Fair, transparent and accountable algorithmic decision-making processes. The premise, the proposed solutions, and the open challenges". *Philosophy & Technology*, vol. 31, núm., 4. DOI:10.1007/S13347-017-0279-X.

LIPTON, Z. C. (2018). «The Mythos of model interpretability». *ACM Queue*, vol. 16, núm. 3 (mayo-junio). Disponible en https://doi.org/10.1145/3236386.3241340 (consultado el 14 de marzo de 2021)

MANCOSU, G. (2019). «Le contentieux des actes pris sur la base d'algorithmes, un point de vue italien». *Revue générale du droit on line*, núm. 49010. Disponible en https://www.revuegeneraledudroit.eu/blog/2019/07/15/le-contentieux-des-actes-pris-sur-la-base-dalgorithmes-un-point-de-vue-italien/ (consultado el 14 de marzo de 2021)

— «Les algorithmes publics déterministes au prisme du cas italien de la mobilité des enseignants». *Rivista Italiana di Informatica e Diritto*, núm. 1. DOI: 10.32091/RIID0005.

MARTÍNEZ HERAS, J. (2020). «¿Cómo aprende la Inteligencia Artificial?». *Guía rápida IArtificial.net*. Disponible en https://www.iartificial.net/como-aprende-la-inteligencia-artificial/#Proceso_del_aprendizaje_supervisado (consultado el 14 de marzo de 2021)

MERCHÁN ARRIBAS, M. (2020). *Guía de uso de la Inteligencia Artificial en el Sector Público*, Colección Tecnologías Emergentes, págs. 6-7. Disponible en https://digitalrevolution.info/guia-de-uso-de-la-inteligencia-artificial-en-el-sector-publico/ (consultado el 14 de marzo de 2021)

MESEGUER YEBRA, J. *et al*. (2017). *Comentarios sobre aspectos clave en materia de acceso a información pública*. Federación Española de Municipios y Provincias y Red de Entidades Locales por la Transparencia y por la Participación Ciudadana. Navarra: Thomson Reuters Aranzadi.

MITTELSTADT, B.; RUSSELL, C.; WACHTER, S. (2019). «Explaining explanations in AI». *Proceedings of FAT* '19: Conference on Fairness, Accountability, and Transparency (FAT* '19)*, January 29–31, Atlanta, GA, USA. ACM, New York, NY, USA, DOI/10.1145/3287560.3287574, Disponible en https://ssrn.com/abstract=3278331 (consultado el 14 de marzo de 2021)

MOLNAR, C. (2018). *Interpretable machine learning. A guide for making black box models explainable*. Leanpub. Disponible en https://christophm.github.io/interpretable-ml-book/ (consultado el 14 de marzo de 2021)

NEW, J.; CASTRO, D. (2018). *How policymakers can foster algorithmic accountability*. Center for Data Innovation. Disponible en https://www2.datainnovation.org/2018-algorithmic-accountability.pdf (consultado el 14 de marzo de 2021)

O'NEIL, C. (2016). *Weapons of math destruction. How Big Data increases inequality and threatens democracy*. New York: Crown.

OSWALD, M. (2018). «Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power». *Philosophical Transactions of the Royal Society A.* Disponible en https://doi.org/10.1098/rsta.2017.0359 (consultado el 14 de marzo de 2021)

OSWALD, M.; GRACE, J.; URWIN, S.; BARNES, G. C. (2018). «Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality». *Information & Communications Technology Law*, Vol. 27, núm. 2. Disponible en https://doi.org/10.1080/13600834.2018.1458455 (consultado el 14 de marzo de 2021)

PEREL, M. y ELKIN-KOREN, N. (2017). «Black box tinkering: beyond disclosure in algorithmic enforcement». *Florida Law Review*, vol. 69, núm. 181, págs. 184-185. Disponible en https://scholarship.law.ufl.edu/flr/vol69/iss1/5/ (consultado el 14 de marzo de 2021)

PONCE SOLÉ, J. (2019). «Inteligencia artificial, Derecho Administrativo y reserva de humanidad: algoritmos y procedimiento administrativo debido tecnológico». *Revista General de Derecho Administrativo*, núm. 50.

RAMIÓ, C. (2018). *Inteligencia artificial y Administración pública. Robots y humanos compartiendo el servicio público*. Madrid: Catarata.

RASHKOVICH, B. (2019). «Government accountability in the age of automation». *Media Freedom & Information Access Clinic*, Yale Law School, 9 abril. Disponible en https://law.yale.edu/mfia/case-disclosed/government-accountability-age-automation (consultado el 14 de marzo de 2021)

RUSTAD, M. L. (2010). *Software licensing. Principles and practical strategies*. Oxford: Oxford University Press.

THE ROYAL SOCIETY (2019). *Explainable AI: the basics. Policy briefing*, pág. 6. https://ec.europa.eu/futurium/en/system/files/ged/ai-and-interpretability-policy-briefing_creative_commons.pdf (consultado el 14 de marzo de 2021)

VALERO TORRIJOS, J. (2018). «La tramitación del procedimiento administrativo por medios electrónicos». ALMEIDA, M.; MÍGUEZ, L. (Dirs.) *La actualización de la administración electrónica*, Santiago de Compostela: Andavira.

VALERO TORRIJOS, J.; FERNÁNDEZ SALMERÓN, M. (Coords.) (2014). *Régimen jurídico de la transparencia en el sector público. Del derecho de acceso a la reutilización de la información*. Navarra: Thomson Reuters Aranzadi.

VIDA FERNÁNDEZ, J. (2018). «Los retos de la regulación de la inteligencia artificial: algunas aportaciones desde la perspectiva europea». DE LA QUADRA-SALCEDO, T.; PIÑAR MAÑAS, J. L. (Dirs.) *Sociedad Digital y Derecho*. Madrid: Boletín Oficial del Estado, Ministerio de Industria, Comercio y Turismo y RED.ES.

VILLANUEVA, J. D. (2020). «Redes neuronales desde cero (I). Introducción». *Guía rápida IArtificial.net*. Disponible en https://www.iartificial.net/como-aprende-la-inteligencia-artificial/#Proceso_del_aprendizaje_supervisado (consultado el 14 de marzo de 2021)

WALDEN, I. (2013). «Open Source as philosophy, methodology and commerce». En SHEMTOV, N.; WALDEN, I.. *Free and Open Software. Policy, Law and Practice*. Oxford: Oxford University Press**.**

WORLD WIDE WEB FOUNDATION (2017). *Algorithmic accountability. Applying the concept to different country contexts*. Disponible en http://webfoundation.org/docs/2017/07/Algorithms_Report_WF.pdf (consultado el 14 de marzo de 2021)