

TRANSPARENCIA Y CONTROL DEL USO DE LA INTELIGENCIA ARTIFICIAL POR LAS ADMINISTRACIONES PÚBLICAS¹

Andrés Boix Palop*

Alba Soriano Arnanz**

1 Este trabajo se ha realizado en el marco de los Proyectos de investigación «Digital economy regulation: equality guarantees provided by public powers and algorithmical tools» (PID2019-108745GB-I00) y «Algorithmical Law» (PROMETEU/2021/009), financiados respectivamente por el Ministerio de Ciencia e Innovación y por la Generalitat Valenciana y “Algorithmic Decisions and the Law: Opening the Black Box” (TED2021-131472A-I00), del Plan de Recuperación, Transformación y Resiliencia. Parte de este trabajo en un formato ensayo ha sido publicada ya por uno de nosotros como “Transparencia en la utilización de inteligencia artificial por parte de la Administración” en el nº100 de *El Cronista del Estado Social y Democrático de Derecho*, dedicado monográficamente al impacto de la Inteligencia Artificial en el Derecho.

* Profesor Titular de Derecho Administrativo, Universitat de València.

** Profesora Ayudante Doctora de Derecho Administrativo, Universitat de València.

I. LAS GARANTÍAS FRENTE A CUALQUIER ACTUACIÓN DE LOS PODERES PÚBLICOS QUE AFECTE A LOS DERECHOS DE LA CIUDADANÍA

El control de la actuación de los poderes públicos es una exigencia consustancial a cualquier Estado de Derecho. Un control que se declina de muy diversos modos, desde garantizando por medio de la división de poderes, en expresión clásica, que “el poder controle al poder”, a limando poco a poco todas las inmunidades de las que tradicionalmente ha disfrutado. Esta lucha contra las atribuciones y potestades extraordinarias y no justificadas del poder público, que entre nuestra doctrina, García de Enterría (aún durante el franquismo) y muy certeramente identificó como una pauta evolutiva de los sistemas jurídicos más avanzados, requiere de muy diversos elementos para que la mezcla funcione (García de Enterría, 1968). Elementos que hay que combinar y que pueden presentarse en diversas dosis, pero que más o menos todas y todos conocemos. Van desde la afirmación de los derechos de la ciudadanía y su profundización al reconocimiento de la plena capacidad de los tribunales de justicia para poder no ya revisar de manera parcial y limitada lo actuado por la Administración sino directamente enjuiciar de modo pleno su corrección jurídica (proceso del que la propia evolución de la justiciabilidad de las actuaciones administrativas y, en el caso concreto de España, de la ampliación paulatina de las efectivas posibilidades de control del poder a cargo de la jurisdicción contencioso-administrativa, hoy consagrada plenamente en el artículo 106.1 de la Constitución –CE–, dan buena cuenta, como estudió con profundidad entre nosotros Fernández Torres [1998]), pasando por la estricta sujeción a la Constitución y resto del ordenamiento jurídico de la actuación de los poderes públicos (art. 103.1 CE).

Un elemento integrante absolutamente esencial de este marco de controles y garantías de la acción del poder público es la publicidad y transparencia de sus actuaciones (Mestre Delgado, 1998 y 2018), que comprende tanto la necesidad de transparencia y derecho de acceso por parte de cualquier ciudadana o ciudadano a la mayor parte de los expedientes administrativos expresada por el art. 105 b) CE (una posibilidad de acceso que ha de ser directamente completa siempre y en todo caso respecto de actuaciones que conciernan singularizadamente a una persona en tanto que interesada como se deduce

del art. 105 c) CE) como también, como no puede ser de otra manera, la propia publicidad de las normas, tanto de las de origen legislativo como de las que son de origen reglamentario, producidas por el propio poder ejecutivo. Esta idea de publicidad de las normas, consustancial a cualquier concepción moderna y democrática del ejercicio del poder, está también constitucionalmente plasmada, y no sólo para las de origen legislativo, en el art. 9.3 CE, donde además queda expresamente enunciada junto a principios como el de legalidad, seguridad jurídica o interdicción de la arbitrariedad de los poderes públicos, sin que sea una casualidad la conexión.

Todas estas garantías, con mejor o peor fortuna y mayor o menor rapidez en lo que se refiere a su implantación en cada país, forman parte del bagaje jurídico que, desde la Ilustración, ha acompañado la evolución de nuestros sistemas jurídicos hacia una concepción avanzada del moderno Estado de Derecho, en su vertiente más objetiva e institucional. En un plano más centrado en los efectos sobre la ciudadanía de la acción pública, además, la actuación de los poderes públicos, y en concreto de la Administración, así enmarcada, ha de ser siempre estrictamente respetuosa con sus derechos fundamentales, que son para todos los poderes públicos generadores directos de obligaciones y deberes para con aquélla por la directa exigibilidad de los mismos y su inmediato despliegue en la esfera jurídico pública. De modo que el principal destinatario de todas estas exigencias, y quien primariamente queda obligado a cumplir con ellas, es el poder público (aunque, por supuesto, los derechos fundamentales también dan lugar a obligaciones de conducta entre particulares, mediante diversos mecanismos y una creciente extensión horizontal de su eficacia, a la que nos solemos referir con el término alemán *Drittwirkung*; los primeros y más importantes destinatarios de las obligaciones que se deducen de ellos son los poderes públicos, que han de quedar sujetos por muchas más obligaciones y sometidos a más y más exigentes controles en su acción por esta razón, derivada de la cautela con que ha de ser tratado el ejercicio de los poderes y potestades exorbitantes que sobre la ciudadanía puede ejercer el poder).

En todo caso, y como es evidente, ninguno de estos avances es lineal ni ha sido sencillo. No son pocas las situaciones en que se han producido retrocesos y, como es consustancial a la tendencia natural del poder a tratar de verse limitado lo menos posible, las formulaciones teóricas o desarrollos prácticos

que han tratado de revertir o minimizar las mejoras sobre el control de la actuación de los poderes públicos y de la Administración han sido también una constante que ha acompañado y todavía acompaña a este mismo proceso de extensión y profundización en estas garantías. Desde las teorías que excluían ciertas categorías de actos del posible control o revisión judicial (por ejemplo, la tradicional y ya abandonada teoría del “acto político”) a tesis que han defendido la inmunidad frente al mismo de las actuaciones discrecionales (o limitado mucho la posible revisión de sus aspectos jurídicamente reglados), pasando por teorías que modulan en ocasiones las garantías cuando se argumenta que la Administración está relacionándose no con ciudadanos o ciudadanas en sentido estricto sino con personas cuya condición difiere de ese estatuto en contextos de “especial sujeción” (por estar algunas categorías de personas, como las y los estudiantes o el personal empleado público, en una situación de mayor dependencia frente a la misma) o que limitan algunas reglas que contienen posibles excesos cuando el ejercicio del poder punitivo del Estado lo protagoniza la Administración y no el poder judicial argumentando su menor lesividad, las doctrinas de muy diverso origen y efectos con el denominador común de restringir el embridamiento a que el Estado de Derecho somete al ejercicio del poder público son habituales. Asimismo, desde un punto de vista más práctico, razones de eficacia, económicas, de despliegue de la acción pública o de mera practicidad de todo tipo son también muchas veces empleadas para cuestionar ciertos controles, su conveniencia o necesidad, en aras a consideraciones ancladas en un supuesto pragmatismo. Esta situación suele producirse, de hecho, siempre que la aparición de novedades tecnológicas obliga a transformar, repensar o traducir al nuevo contexto técnico algunas de las garantías tradicionales por las que tanto ha costado luchar y que finalmente han quedado asentadas por el convencimiento de la comunidad jurídica y de la ciudadanía de su necesidad. Nada tan cómodo como apelar a que los cambios económicos, sociales o tecnológicos han modificado algunos de los equilibrios que las hacían necesarias, o han conllevado nuevos problemas para su puesta en marcha que las hacen desaconsejables en toda su extensión, para acabar introduciendo por la puerta de atrás contrarreformas nada garantistas que, además, las más de las veces, acaban por demostrarse a la larga no sólo insuficientes sino generadoras de no pocos problemas.

El creciente empleo de la Inteligencia Artificial (IA) en diversas fases de la actuación de los poderes públicos, por medio de algoritmos o programas

que incluyan la predeterminación bien de la evaluación de las circunstancias que han de ser tenidas en cuenta por la Administración pública, bien que lleguen directamente por sí mismos a la solución (o ayuden en estas tareas), quizás todavía incipiente hoy en España (aunque esta afirmación puede y debe ser matizada cada día un poco más), es un ejemplo claro de una circunstancia que puede dar pie a que aparezca de nuevo la tentación de excluir o minimizar la aplicación de algunas garantías (Barona Vilar, 2022; Martínez Garay y García Ortiz, 2022). Por ejemplo, por no ser cómodo, o no ser sencillo, ni barato, ni a veces siquiera evidente, cómo deban trasladarse algunos de estos controles tradicionales cuando se emplean sistemas de inteligencia artificial puede caer en la tentación de, directamente, prescindir de algunos de ellos. A fin de cuentas, se dirá en ocasiones, el propio avance tecnológico hace innecesarias tantas cautelas. O, aunque no sea así, aporta tantos beneficios que impedir o ralentizar su desarrollo con consideraciones garantistas sería muy negativo en términos de eficiencia y por ello indeseable socialmente.

Sin embargo, y como criterio general sobre el que es bastante sencillo ponerse de acuerdo, hay que dejar claro que una consecuencia de este tipo derivada de la implantación de sistemas de IA en el quehacer administrativo sería de todo punto inaceptable: el despliegue tecnológico que la evolución de la informática puede aportar al ejercicio del poder público y la actividad administrativa hay que exigir que se lleve a cabo a la par que se conservan todos los controles y garantías respecto de la actuación administrativa tradicionales debidamente adaptados, traducidos, transformados, en lo que sea necesario, a la nueva realidad (Cerrillo i Martínez, 2019). Aunque, eso sí, garantizando que siguen en vigor en la medida en que son la efectiva plasmación de la necesidad de proteger materialmente ciertos ámbitos y esferas de autodeterminación y protección de la ciudadanía que subsisten, también cuando en la acción del poder ejecutivo se empleen algoritmos, programas o cualquier forma de inteligencia artificial, sin que el recurso a estas herramientas pueda suponer en ningún caso una justificación de la relajación de los mismos o la aparición de “nuevas inmunidades del poder” (Martín Delgado, 2009; Valero Torrijos, 2019).

En definitiva, y como idea absolutamente básica, la ciudadanía debe conservar todos sus derechos frente a la actuación administrativa también cuando

ésta es realizada por medios algorítmicos o empleando herramientas de IA (Cotino Hueso, 2020; Presno Linera, 2022). El equilibrio entre la protección jurídica de su posición jurídica y la capacidad de actuación del Estado, del poder, no ha de verse alterado, y en ningún caso menguadas las garantías, por el simple hecho de que los poderes públicos dispongan de herramientas técnicas que les permitan ser más eficientes en su actuación. Es más, en todo caso, lo contrario sería lo correcto: la alteración del equilibrio preexistente razonable, si ha de darse, debiera ser en términos de proteger más a los ciudadanos si las herramientas de que dispone el poder se convierten en más poderosas.

El problema de esta declaración genérica, por mucho que resulte sencillo ponerse de acuerdo en abstracto sobre ella, es su concreción, esto es, alcanzar un mínimo consenso sobre qué ha de suponer a la hora de regular los efectos y usos de la IA en los quehaceres administrativos en punto a cómo trasladar a estos supuestos la protección tradicional que nuestros ordenamientos jurídicos han ido decantando para los ciudadanos. Este paso es complicado y, por lo general, suele llevar un tiempo, tras los inevitables balbuceos iniciales, realizarlo de modo satisfactorio. De hecho, y al menos hasta la fecha, razones de eficiencia, económicas o de mera dificultad de aprehensión de todas las consecuencias que suponen las transformaciones de la tecnología empleada por los poderes públicos han llevado a que la mayor parte de usos administrativos que por el momento se realizan en España, así como en otros países de nuestro entorno, empleando sistemas de IA en la actuación administrativa, adolezcan de una llamativa falta de control respecto de algunos de sus aspectos. Llamativa no sólo porque es sorprendente constatar hasta qué punto el rearme tecnológico, cuando empleado por los poderes públicos, ha conducido con naturalidad a eliminar garantías para la ciudadanía sin que hayan sonado inmediatamente todas las alarmas. Es más sorprendente todavía si tenemos en cuenta que cuando estas mismas herramientas son empleadas por actores privados, como es el caso de agentes económicos del estilo de las cada vez más omnipresentes plataformas digitales, la reacción del Derecho, tanto europeo (véase la reciente *Digital Markets Act* que impone obligaciones de cierta exigencia a las grandes plataformas cuando emplean medios tecnológicos que puedan permitirles ciertos privilegios en sus interacciones de mercado) como español (es el caso de la conocida como “ley rider” que trata de garantizar derechos

a los trabajadores de algunas de estas plataformas) ha sido mucho más enérgica, imponiendo obligaciones y estableciendo correlativos derechos de la ciudadanía frente al uso de algoritmos o de IA que son mayores que las que el ordenamiento jurídico español y europeo prevé para los casos en que son los poderes públicos los que emplean esta tecnología (en un llamativo caso de lo que Goerlich Peset denomina habitualmente en sus intervenciones en público sobre estos temas “*Drittwirkung* inversa” (Goerlich Peset, 2022), que a la postre lo que demuestra es que, en efecto, el poder tiene una inveterada y constatable aversión a autolimitarse, incluso en supuestos estrictamente equivalentes a casos donde en el mundo jurídico-privado el poder público no ha tenido duda ninguna en la necesidad de proteger la posición jurídica de los ciudadanos que estaban en esa misma posición).

Es ésta una situación que urge revertir cuanto antes, máxime cuando el empleo de estas herramientas está llamado a ir a más en el futuro, so pena de que nos enfrentemos a situaciones donde la posición jurídica de los ciudadanos pueda quedar cada vez más comprometida y en grave retroceso en términos de garantías respecto del punto de llegada actual, logrado tras dos siglos de decantación y lucha por la mejora de los controles y reducción de las inmunidades del poder.

Un ejemplo paradigmático de esta situación, y quizás aquél sobre el que más se ha reflexionado hasta la fecha, es el de la falta de transparencia y conocimiento sobre el funcionamiento de ciertos algoritmos, en algunos casos (que son en realidad, desgraciadamente, casi todos), cuando son empleados por la Administración pública para tomar decisiones (o para ayudar a la adopción de las mismas de forma suficientemente relevante como para entender que pueden predeterminarla o condicionarla). Una falta de transparencia que, más allá de si es compatible o no con lo dispuesto en el art. 9.3 CE, dificulta gravemente la posibilidad efectiva de una verdadera auditabilidad sobre el funcionamiento de estas herramientas, lo que junto a los problemas que puedan derivarse de poner en cuestión la posibilidad efectiva del debido control sobre el rigor jurídico y no afección a los derechos de la ciudadanía del proceso de toma de decisiones, puede suponer una grave quiebra de la positiva dinámica de lucha contra las inmunidades del poder que en nuestro país y tantos otros ha caracterizado, para bien, la evolución del Derecho público en las últimas décadas.

II. EL PROBLEMA

II.1. El uso de programas y algoritmos con IA por las Administraciones Públicas para la mejora de su actuación

Como ya se ha apuntado, y es además bien sabido, el uso de programas informáticos de procesamiento de datos en los procesos de toma de decisiones públicas y privadas no es en realidad un fenómeno nuevo. Especialmente, cuando este uso es como herramienta de apoyo. De hecho, nuestras Administraciones Públicas emplean desde hace años sistemas informáticos, de los más sencillos a algunos crecientemente complejos, para almacenar y cruzar información. La novedad no viene tanto de esta situación como de que cada vez nos encontramos con más supuestos en los que se emplean programas y algoritmos que, con un apoyo cada vez más generoso en sistemas de IA, e incluso con el recurso a modelos de aprendizaje profundo automatizado (*machine y deep learning*), que pueden superar la capacidad de evaluación de las situaciones que tiene la propia Administración a partir de sus herramientas tradicionales o, incluso, derivar de esa evaluación más precisa de la situación la adopción de criterios o decisiones más ajustadas a lo querido por la norma o principios jurídicos básicos que regulan la actividad de que se trate, de modo que poco a poco los resultados de esta evaluación o de la propuesta de pautas decisorias derivadas de la ejecución de los programas o algoritmos acaben sustituyendo al proceso de toma de decisiones por medio de inteligencias humanas al que estamos acostumbrados y sobre el que se ha construido nuestro Derecho (y sus correlativos controles).

Sabemos que no hablamos de ciencia ficción ni de posibilidades futuras, sino del aquí y del ahora, incluso en países como España que no es en esta materia, por lo general, punta de lanza de desarrollos tecnológicos y donde solemos emplearlos cuando ya han sido desarrollados, probados y asentados en otros lugares. Así, por poner el que es el ejemplo más conocido en nuestro país, el Ministerio del Interior viene desarrollando desde hace más de una década un sistema informático (VioGén) que, a partir de los datos con que se cuenta en cada caso denunciado de violencia de género, hace una evaluación del riesgo potencial que tienen las víctimas de volver a sufrir una agresión caso por caso. Esta evaluación, por lo general, se considera ya muy fiable, hasta el punto de poder ser más ajustada que la que haría

un especialista medio, por lo que todos los agentes implicados en proteger a las víctimas posibles la tienen muy en cuenta, desde la propia policía a jueces y fiscales. La herramienta, aunque no contiene hasta la fecha el empleo de IA en el sentido de que el propio algoritmo pueda ir aprendiendo de los datos introducidos y alterando las ponderaciones con las que ha sido programado, sustituyéndose así a la inteligencia humana, ya anuncia su posible inclusión en los nuevos desarrollos en curso para su mejora. Por su parte, y aunque de nuevo es un sistema en el que la ponderación de lo que cuenta cada elemento que calculará el programa por medio de un conjunto de algoritmos está predeterminada por los programadores sin que haya un proceso de *machine learning* en el mismo, la Generalitat de Catalunya emplea para la clasificación de sus presos y la evaluación de su conducta a efectos de la gestión penitenciaria e incluso de las políticas de reinserción un sistema (Riscanvi) que también es esencialmente algorítmico y al que, por la objetividad y seguridad que confiere de trato aparentemente neutro, se concede una creciente importancia (Martínez Garay y Montes Suay, 2018; Martínez Garay y García Ortiz, 2022). Ambos sistemas recuerdan a los empleados en países como Estados Unidos para calcular riesgo de reincidencia, que han dado lugar a alguna de las discusiones jurídicas más encendidas en los últimos años sobre la admisibilidad de la aplicación de estos programas, que pueden condicionar de modo muy notable decisiones que afectan a la libertad personal, cuando no se conoce exactamente su funcionamiento ni la ponderación exacta que realizan, por tratarse de *software* de apoyo (además, propiedad de las empresas desarrolladoras) para la adopción de decisiones públicas en ejercicio de autoridad (caso Loomis, en Estados Unidos, sobre el que ha escrito de manera informada y certera en nuestro país Martínez Garay [2018]).

En un plano quizás aparentemente menos dramático en términos de libertades, pero más cotidiano para millones de ciudadanas y ciudadanos, la Inspección de la Seguridad Social (Sánchez Galán *et al.*, 2019), la Agencia Española de Administración Tributaria (Rodríguez Peña, 2021) o la Comisión Nacional de Mercados y Competencia (Campuzano Fernández y Vega Vicente, 2021) emplean ya sistemas algorítmicos para la detección del fraude con toda normalidad, y desde hace años, como es el caso también de sus equivalentes en todos los países de nuestro entorno (aunque, sorprendentemente, sin que la ciudadanía española sepa exactamente qué programas y en qué situaciones se

emplean, pues con el argumento de que son sólo herramientas de apoyo a la decisión última, pero que ésta siempre depende de una decisión humana posterior, se entiende que no es necesario proporcionar esta información). Lo hacen a partir de la recopilación y cruce de datos, como se hace en todo el mundo tanto en el sector público como en el privado, para lo que los algoritmos, y especialmente los que emplean IA para establecer correlaciones y aprender de las mismas detectando patrones, son de mucha utilidad. Además, no sólo estas actividades de supervisión y control, sino también la gestión y optimización de muchas políticas públicas, desde la atribución de viviendas públicas a las ayudas sociales, son en cada vez más países supervisadas, controladas, cuando no directamente determinadas, por programas informáticos, con un mayor o menor uso de la IA. En el caso español, por ejemplo, la gran batalla jurídica producida hasta la fecha en la lucha por lograr la transparencia algorítmica, liderada por la asociación Civio, tiene que ver con una política social: en concreto, la asignación del bono eléctrico a las y los ciudadanos más vulnerables para ayudarles a hacer frente al coste de la energía, que aunque se realiza a partir de los criterios establecidos reglamentariamente, es declinada y concretada por un algoritmo (el programa en cuestión recibe el nombre de BOSCO) que determina en la práctica a quién se concede el bono y a quién no. Frente a las dudas que generaban los resultados del algoritmo, la asociación ha solicitado sin éxito, primero en vía administrativa, luego ante el Consejo de Transparencia y finalmente en sede judicial el acceso a la codificación del programa, a fin de poder entender su funcionamiento y las razones por las que aparentemente en ocasiones los resultados no parecían ser coherentes con las normas reglamentarias que determinan los criterios de concesión, a efectos de poder garantizar un mejor control de posibles errores que ayude a defender los derechos de las posibles personas afectadas (De la Cueva, 2019). La sentencia del caso BOSCO, de momento de instancia (Sentencia del Juzgado Central de lo Contencioso Administrativo nº 8 de la Audiencia Nacional de 30 de diciembre de 2021) y recurrida ante el Tribunal Supremo por Civio, reconoce, sorprendentemente, el derecho de la Administración a no compartir el código fuente del algoritmo empleado apoyándose en algunas de las razones que tradicionalmente han sido aducidas para ello, aunque sin hacer hincapié en la naturaleza propietaria del *software* en cuestión y su propiedad intelectual: el hecho de que en realidad el programa sea simplemente una herramienta de apoyo para determinar a quién se concede o no la ayuda, por mucho que directamente determinante, dado que luego habrá una decisión de una

persona que asume esta evaluación y que podrá ser en todo caso recurrida en cada caso, así como supuestas razones de seguridad que se sostiene que podrían provocar que un acceso completo al código fuente del algoritmo pudiera comportar riesgos de seguridad para la estabilidad de la plataforma (que no se explicitan tampoco en exceso). Como puede constatarse, ambas razones son más bien endeble: la posibilidad de recurso en cada caso, ¡qué menos!, no excluye que estos puedan y deban ser acompañados de conocer cómo funciona el algoritmo, tanto a efectos de facilitar las cosas a las y los ciudadanos concretos en sus recursos como para un mejor control general del funcionamiento de la herramienta; los riesgos de seguridad por acceso a un algoritmo de esta naturaleza debieran ser inexistentes y, si efectivamente se pudieran dar, es responsabilidad de la Administración evitarlos por otros medios que no supongan eliminar la posibilidad de acceder a conocer cómo se realiza en la práctica el cálculo.

Estos mismos problemas, como es evidente, no se producen sólo en España, y estas mismas discusiones y dudas se están generando en todo nuestro entorno comparado allí donde se emplean programas o algoritmos, y más todavía si introducen modelos de IA más ambiciosos, para la mejora de la actuación administrativa. La tendencia a la opacidad es también común, con casos que pueden calificarse de escandalosos, como el ya referido Loomis en los Estados Unidos (Martínez Garay, 2018), pero también con los primeros ejemplos de una reacción jurídica frente a esta situación. Quizás la más conocida, por la importancia de la apuesta que había realizado el gobierno de los Países Bajos con la puesta en marcha de la herramienta, es la Sentencia de la Corte de Distrito de La Haya de 5 de febrero de 2020, que vetó la aplicación de un programa que pretendía hacer uso de algoritmos que podían detectar pautas y establecer correlaciones para la asignación y control de ayudas sociales, el denominado Sistema de Indicación de Riesgos (SyRI), por entender que sólo será lícito emplear este tipo de instrumentos siempre que haya un claro interés público que supere con creces los riesgos de afección a la intimidad de las personas y que se tomen las medidas adecuadas para que la aplicación del sistema sea llevado a cabo sin que se generen riesgos de sesgos, discriminaciones o estigmatización de los colectivos más vulnerables. Para ello, y en ausencia de normas europeas o nacionales que obliguen a desvelar el código fuente, permitiendo así un mejor control sobre el mismo, o que restrinjan estos usos de la informática en caso de que

pueda sospecharse la existencia de estos riesgos, la Corte directamente acude al Convenio Europeo de Derechos Humanos (CEDH) y a la protección que el mismo hace sobre la intimidad de las personas para fundamentar la prohibición del uso de la herramienta (Cotino Hueso, 2020). Esta sentencia es de una gran importancia e interés y nos demuestra hasta qué punto la cuestión preocupa, y no sólo en relación con la protección de la intimidad, sino también por la necesaria prevención de generación de los mencionados sesgos, pero sobre todo por cómo permite constatar que no disponemos aún de marco jurídico claro que delimite cómo y en qué condiciones se puede hacer uso de estos sistemas algorítmicos, razón por la cual se ha de recurrir a una ponderación con base en el CEDH para declarar antijurídica la herramienta en esas condiciones de opacidad (y por ello que no hay una suficiente justificación de la necesidad de su uso). Como puede verse, el problema no es sólo español y la tendencia a la falta de transparencia, así como la necesidad de combatirla, es consustancial a la aplicación de estos sistemas algorítmicos en todos los ordenamientos jurídicos.

II.2. Riesgos derivados del funcionamiento técnico de los algoritmos y de la IA empleados por las Administraciones Públicas

Como es evidente, este cambio (y los posibles problemas asociados al mismo) será tanto mayor cuanto estos algoritmos o programas cuenten con más autonomía y sus resultados no vengan estricta y completamente pre-determinados por las personas encargadas de diseñarlos sino que éstas se limiten a alimentar los programas con grandes cantidades de datos para entrenarlos, proporcionándoles una serie de instrucciones más o menos extensas pero que, en todo caso, dejen un margen de apreciación al propio sistema, a través de sistemas de aprendizaje autónomo y, sobre todo, donde sea la propia máquina la que determine cómo se concreta la aplicación de unas instrucciones que más que reglas serían, más bien, una suerte de principios, de mandatos de optimización, sobre lo que se desea conseguir. Es decir, los riesgos y la necesidad de una clarificación jurídica son tanto mayores cuanto mayor sea el uso de IA avanzada por los programas en cuestión, debiendo estudiarse cuáles son los posibles riesgos concretos del empleo de IA que pueden superponerse a las preocupaciones tradicionales

en cuanto al control y garantía de la justicia y legalidad de las decisiones administrativas en cuestión.

A estas alturas ya sabemos que junto a las ventajas que conlleva el uso, mejora y generalización de estos instrumentos, y de ahí su indudable éxito y que su expansión y generalización paulatina en el futuro sean inevitables, es evidente que los sistemas algorítmicos que evalúan y tratan de predecir la realidad social y que son empleados en contextos de toma de decisiones plantean riesgos específicos. Unos riesgos que, proyectados sobre los quehaceres administrativos y, muy específicamente, sobre el ejercicio de atribuciones de autoridad por parte de los poderes públicos que impacten sobre la esfera de derechos de los ciudadanos, obligan a un cuidado y unas garantías adicionales por los efectos que pueden tener. Sobre éstos hay ya una copiosa bibliografía, que ha identificado los más importantes y habituales, así como los más potencialmente graves desde un punto de vista jurídico por su afección a los valores y principios sobre los que se asientan los Estados democráticos y los derechos fundamentales de la ciudadanía (ver, sobre esta cuestión, Soriano Arnanz, 2021a). Aunque no podemos revisarlos ahora con detalle, pues ello nos desviaría del objetivo de este trabajo, sí conviene, al menos, tenerlos presentes:

- *Singularidad, autonomía y protección de datos personales.* Los sistemas automatizados de toma de decisiones son capaces de generar perfiles muy precisos, pero lo hacen considerando la pertenencia de la persona en cuestión a diferentes grupos. Es decir, no se consideran las especificidades y singularidad propias de cada uno de los seres humanos que el sistema analiza, sino que la representación virtual de cada ciudadano es el producto de su posición con respecto a otros individuos y de la ubicación que de ello determina el programa respecto de su supuesta pertenencia a grupos de individuos que compartirían características que, por ello, se le van a presuponer. Esto puede conducir a la toma de decisiones erróneas ya que lo que es verdad para el grupo, no tiene por qué serlo necesariamente para una persona en concreto (Zarsky, 2013 y 2016). Directamente vinculado a esta cuestión, hay un importante riesgo de que la información así tratada y construida facilite bien un gran control, bien directamente ciertas formas de manipulación de los ciudadanos así definidos (O’Neil, 2016: 70). Junto a ello, los daños producidos por las tecnologías de procesamiento de datos

que probablemente resultan más evidentes son los que se generan en el ámbito de la intimidad o privacidad de las personas, que puede verse violentada de manera desproporcionada e injustificada. Es precisamente el caso del sistema SyRI ya comentado, y justamente ésta la razón por la que fue descartado (Cotino Hueso, 2020).

- *Sesgos, errores y discriminación.* Cabe también destacar el riesgo de que los resultados ofrecidos por estos sistemas, y que son luego empleados en la toma de decisiones que afectan a la ciudadanía, puedan ser total o parcialmente erróneos o estén sesgados (y, sobre todo, que el empleo de herramientas algorítmicas pueda provocar que estos problemas se retroalimenten y pasen más inadvertidos, y sean más difíciles de detectar y solucionar, precisamente por el tipo de instrumento que los genera, como ha desarrollado de forma muy detallada O’Neil, 2016). Sesgos que, además, como se ha constatado en los últimos años, tienden a repetir y perpetuar las dinámicas ya asentadas de dominación, privilegio y discriminación, lo que los hace particularmente odiosos porque pueden introducir un riesgo adicional de que éstas queden acentuadas y reforzadas. Como es evidente, estas posibles discriminaciones y sesgos suelen afectar más a las personas en una situación socioeconómica más vulnerable (Ranchordas y Schuurmans, 2020) o a grupos tradicionalmente discriminados como las mujeres, debido a que los sesgos de género subyacentes a la discriminación que padecen en muchos ámbitos tenderá a reproducirse cuando se emplean estos sistemas algorítmicos de evaluación de la situación si no se ponen medidas que lo eviten, lo que requiere de una intensa y profunda evaluación inicial, así como del establecimiento de medidas de evaluación del funcionamiento del sistema para su posible corrección (Soriano Arnanz y Simó Soler, 2021). En este sentido, cabe destacar que las situaciones de discriminación y perpetuación de la desigualdad derivadas del uso de herramientas de IA pueden venir producidas tanto por entrenarse el sistema con una base de datos sesgada (por ser la base de datos el producto de decisiones históricas –sesgadas y discriminatorias– tomadas por seres humanos, algo particularmente habitual en el caso de los sistemas empleados en el contexto de la selección de recursos humanos) como por el *feedback* recibido por el sistema una vez este se pone en marcha. Así, por ejemplo, en este segundo caso, la perpetuación de estereotipos negativos asociados a grupos históricamente desaventajados por parte de los motores de búsqueda de Internet (por ejemplo cuando

determinadas combinaciones de palabras referentes a mujeres racializadas generan resultados en los que aparecen una gran cantidad de imágenes que hipersexualizan a aquéllas) son, en muchos casos, el resultado de las preferencias de las personas usuarias de la red. Asimismo, en otras ocasiones, el problema no es que el sistema esté sesgado por los datos de entrenamiento empleados o que aprenda a ofrecer resultados sesgados una vez puesto en marcha, sino que las situaciones de discriminación o perpetuación de la desigualdad son el resultado de las decisiones tomadas por la organización que empleará el sistema (Soriano Arnanz, 2021b). Esto puede verse de forma meridiana en los casos en los que las Administraciones deciden emplear estas herramientas con el objetivo de detectar el fraude entre las personas receptoras de ayudas, situándolas así en una posición de mayor vulnerabilidad (Ranchordás y Schuurmans, 2020), o en casos en los que se prioriza el ahorro y la eficiencia económica sobre la protección de las personas en situaciones de desventaja, como en el caso del algoritmo que durante un tiempo fue empleado para distribuir recursos entre las personas desempleadas en Austria (Soriano Arnanz, 2022).

- *Control privado, y no público, del código y de los datos.* Las situaciones descritas en los dos párrafos anteriores no serían tan problemáticas si las Administraciones públicas pudieran ejercer un control efectivo sobre los sistemas algorítmicos adquiridos o empleados para ayudar a la toma de decisiones administrativas, pero no suele ser el caso. Hasta la fecha, en casi todo nuestro entorno comparado y por supuesto también en España, la Administración no dispone de los medios y capacitación tecnológica, y a veces ni siquiera de la voluntad, para supervisar y controlar los sistemas automatizados empleados en estos contextos. La inexistencia de obligaciones jurídicas respecto del control sobre los algoritmos o programas hace que, en la mayor parte de ocasiones, estas soluciones informáticas de apoyo se contraten para ser empleadas, aceptando las condiciones de uso impuestas por los desarrolladores de manera acrítica. Una de ellas suele ser, además, la no publicación del código fuente (aunque es cierto que este problema no es único de estos casos, pues desarrollos propios, como el ya comentado VioGén, que sí es una creación de la propia Administración española, comparten esta muy criticable tendencia a la opacidad). En cualquier caso, junto a este problema, el no control sobre el *software* empleado supone un evidente conflicto potencial de intereses con no pocos riesgos, por cuanto

se emplea una herramienta que puede afectar a derechos de la ciudadanía el control de la cual no se tiene. En ocasiones, incluso, encontramos casos donde las empresas privadas que crean estos sistemas llegan a licenciar su uso a las Administraciones públicas a cambio de datos que obran en poder de éstas, en lo que es un supuesto extremo de malas prácticas.

- *Opacidad e imposibilidad de acceso al código fuente del algoritmo.* La dificultad de acceder a la lógica subyacente a las decisiones tomadas por sistemas automatizados y los problemas generales de transparencia que caracterizan la utilización de estos sistemas complica enormemente que las personas que se vean afectadas por el uso de éstos puedan defender sus derechos e intereses legítimos frente a las posibles agresiones provocadas por su uso. De hecho, en muchos casos porque los ciudadanos ni siquiera saben que están siendo objeto de una evaluación algorítmica, por ejemplo, por empresas privadas (pero también de tratamientos públicos cuando, alegando que son meras herramientas de apoyo, ni siquiera se informa de ello a los ciudadanos, como es el caso por ejemplo de los algoritmos empleados para la inspección en España). Pero, como es obvio, resulta particularmente peligroso cuando se trata de programas que se encargan de la toma de decisiones públicas automatizadas (Orofino, 2020), por cuanto el ejercicio de autoridad y las decisiones de los poderes públicos impactan aún más, y con más intensidad, sobre la esfera de derechos y deberes de los ciudadanos. Por ello, los controles y exigencias de transparencia deberían ser mayores para los algoritmos empleados por el sector público, a fin de garantizar el estricto respeto por parte de la Administración de los derechos de los ciudadanos (Vestri, 2021). Algo que no siempre ocurre, y ello a pesar de que el principio y mandato de transparencia desempeña un papel fundamental en los ordenamientos jurídico-administrativos de cualquier Estado democrático (Mestre Delgado, 1998 y 2018). Adicionalmente, la propia función de un algoritmo o programa que emplea IA para ayudar a la toma de decisiones públicas, o incluso cuando es el programa el que directamente acaba adoptándola debería resultar un elemento a la postre determinante para garantizar una total transparencia, combatiendo así este riesgo y, de paso, ayudando a dar una mejor respuesta jurídica, también, a los anteriores. Es, por esta razón, absolutamente clave.

III.3. ¿Han de ser tratados los algoritmos que predeterminan decisiones administrativas como reglamentos porque funcionalmente actúan como tales y, por ello, hemos de adosarles las garantías jurídicas de publicidad y control tradicionalmente predicadas de los mismos?

En cualquier caso, y respecto de todos estos riesgos, sea mayor o menor el grado de predeterminación del algoritmo o programa, lo que podemos constatar es que estos instrumentos, cuando empleados para ayudar a la adopción de decisiones administrativas, están condicionando o incluso directamente estableciendo cuál vaya a ser la decisión final, en la medida en que ésta vaya a depender, total o parcialmente, de esa evaluación. Y, para ello, actúan a partir de una programación, de un código informático, que determina exactamente en qué términos se ha de producir esta operación. En trabajos previos ya hemos argumentado que esta función, respecto de los resultados derivados de la aplicación de algoritmos, es estrictamente equivalente a la que los reglamentos realizaban tradicionalmente, y todavía realizan en gran parte del quehacer administrativo actual, respecto de las decisiones administrativas (ver, sobre todo, Boix Palop, 2020).

Si estos efectos prácticos son equivalentes, por no decir los mismos, habría que empezar a asumir de una vez con todas las consecuencias que, al menos en estos casos, el *código fuente también es código jurídico*. Lo es, de hecho, a todos los efectos. Lo es porque ese código delimita el efectivo marco de actuación de los particulares, marca los límites a sus derechos y genera una producción con efectos jurídicos que fácilmente podemos llamar “actividad administrativa” automatizada a partir de esos parámetros previamente configurados; estamos pues en estos casos ante algo muy parecido a normas reglamentarias. Y lo es también porque el código fuente es el reglamento que preordena los márgenes de la actuación de la *inteligencia artificial* que, más o menos modesta según los casos, ejecuta sus instrucciones de un modo conceptualmente parangonable, de forma estricta además, a la actuación que nuestro Derecho presupone a una inteligencia humana convencional cuando ha de ejecutar una actuación administrativa normativamente enmarcada.

En definitiva, si los “algoritmos empleados por la Administración pública de modo no puramente instrumental producen materialmente los mismos efectos que cualquier reglamento (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) sería bueno entenderlos como reglamentos en un sentido jurídico material y así tratarlos por el Derecho a la hora de regular cómo se producen, y aplican, las garantías en torno a estos procesos (Boix Palop, 2020)”. De este modo la consecuencia jurídica inmediata sería la necesidad de traslación directa de todo el régimen jurídico que enmarca y controla el ejercicio de la potestad reglamentaria (debidamente adaptado y matizado, si se quiere) para estos usos, lo que redundaría en que nos encontraríamos ya con un modelo de garantías preconfigurado perfectamente útil para aplicar en este campo. Hay quien ha criticado esta conclusión sobre la naturaleza jurídica de los algoritmos que encuadran y predeterminan el ejercicio de funciones administrativas o conforman el código general a partir del cual se concretarán las decisiones concretas (Huergo Lora, 2021), por considerar que la analogía no es exacta o que se puede incluso pensar en clave de acto administrativo antes que de reglamento para encuadrar jurídicamente estas herramientas. En lo que hay un acuerdo sin embargo general es en la necesidad de aplicar todas las garantías al control de esta actividad, aunque en ocasiones ello haya de hacerse con la debida “traducción” a un contexto tecnológico que funciona de modo diferente.

Ahora bien, si analizamos con un mínimo de exigencia cuáles habrían de ser esas garantías mínimas, resulta que son estrictamente equivalentes a las que, más o menos, hemos ido decantando arduamente durante décadas como absolutamente necesarias en un Estado de Derecho para cualquier norma reglamentaria. Lo cual no es llamativo, pues, como se ha dicho, plantean materialmente los mismos problemas. Yendo un poco más allá en este punto, un régimen mínimo de garantías para la elaboración y uso de algoritmos por parte de las Administraciones Públicas que se derivara de su consideración como norma reglamentaria supondría aplicar a los mismos, al menos, las siguientes pautas (todas ellas perfectamente razonables, por mucho que puedan necesitar de ajustes en su concreta traducción o traslación al control de programas y algoritmos empleados para adoptar decisiones públicas):

— *Reglas referidas al proceso de elaboración de estos códigos algorítmicos para su uso administrativo que garantizarían que se desarrollaran de una manera estrictamente pautada, con información pública, participación ciudadana, etc.* Se puede plantear, por ejemplo, aplicarles si no todo el nuevo Título VI de la Ley de Procedimiento

Administrativo Común (LPAC) y sus normas de desarrollo respecto de cómo han de elaborarse las disposiciones reglamentarias y los procedimientos para ello, sí garantizar que los principios de la misma, que buscan un efectivo conocimiento de la ciudadanía sobre cuáles son las pretensiones de la regulación y cómo se plantea llevarlas a cabo, se vean cumplidos de una manera que garantice esta información y participación en el proceso de elaboración (en este mismo sentido habría que establecer reglas diferenciadas para los ajustes y modificaciones menores, que en materia de programas y algoritmos son constantes, y que estas modificaciones que no alteran los objetivos sino que son mera actualización técnica, no debieran pasar por todo el proceso). Sólo de este modo puede lograrse que la producción de las normas que *de facto* van a gobernar cada vez más las relaciones interadministrativas y a enmarcar los derechos de la ciudadanía sean no sólo de la suficiente calidad técnica (pero también jurídica aunque estemos hablando aparentemente de la mera programación informática) sino que además pueda desarrollarse desde su origen un control adecuado sobre el funcionamiento de las aplicaciones y de los diversos sistemas automatizados.

- *En lógica correspondencia con esta necesidad de control, y una vez aprobada, la programación correspondiente, así como todo su código fuente, debieran ser, al menos tendencialmente y siempre respecto de sus elementos esenciales, públicos* (del mismo modo que lo son, y de nuevo aparece una evidente identidad de razón, las normas reglamentarias por imperativo constitucional ex art. 9.3 CE). Toda la ciudadanía debiera poder conocer y revisar hasta el último detalle de esta programación, al menos en todos aquellos casos y en la medida en que tengan que ver con el ejercicio de sus derechos, pero también con carácter general de modo que se pueda entender por cualquiera con los conocimientos y el tiempo necesarios el exacto funcionamiento del código y que se puedan desentrañar tanto los efectos del mismo como, en su caso, sus posibles fallos y defectos. Como veremos posteriormente, al desarrollar los distintos modelos de auditabilidad algorítmica habituales hasta la fecha, estamos muy lejos de haber logrado que esta garantía aparentemente tan evidente y necesaria haya sido asumida por nuestras Administraciones Públicas.
- Además, esta programación, reforzando la idea de que estamos ante una realidad con una clara identidad de razón con cualquier norma reglamentaria,

debiera poder ser cuestionada y analizada en cada supuesto aplicativo concreto, y tras la constatación en su caso de la existencia de defectos en la misma o de confirmarse que su funcionamiento limita o cercena derechos de la ciudadanía, debiera poder ser impugnada por ello, con independencia de cuándo hubiera sido aprobada. De nuevo vemos que, más allá de la cuestión puramente terminológica y conceptual, las garantías materiales por las que vela el establecimiento de la posibilidad de un procedimiento de recurso tanto directo como indirecto contra reglamentos tiene toda la lógica que sean aquí replicadas porque los problemas prácticos y riesgos concretos que puede suponer una programación informática fallida, un código fuente erróneo, son los mismos que los que supone un reglamento ilegal. De hecho, en sentencias como la del caso BOSCO ya citadas, al menos, se tiene claro que esta garantía ha de ser en todo caso respetada (ahora, bien, se reconoce la necesidad de respetar la posibilidad de que en todo caso pueda recurrirse la decisión denegatoria caso por caso, pero sin referirse a lo que podría ser el recurso indirecto que de ello podría y debería derivarse contra el algoritmo en sí, caso de que quede acreditado que el problema está en la herramienta, por no adecuarse correctamente a lo que determina el marco jurídico de acuerdo al cual habría de, teóricamente, operar).

Obviamente, conseguir llegar a una situación en que se respeten estas garantías también para los algoritmos empleados por la Administración es difícil. Pero no por ello ha de dejar de ser exigido. En este sentido, ha de criticarse la posición que califica de ilusoria la pretensión de controlar por medio del Derecho cada uno de los detalles y aspectos concretos del funcionamiento de las aplicaciones informáticas y los sistemas de información, por antieconómica o directamente imposible, por cuanto no se trata de ese control absoluto, sino de la garantía del cumplimiento de los principios señalados. Y por mucho que ello sea difícil y exigente, es a lo que hemos de tender (como ha resaltado, por ejemplo, Valero Torrijos [2019]). En la medida en que nos refiramos a las aplicaciones y sistemas que está empleando la Administración pública, y más aún cuando se trata de las que ésta emplea en sus relaciones con los ciudadanos, esta aspiración no sólo no es ilusoria sino que debiera ser absolutamente esencial. Y con todo detalle y rigor, es decir, llegando a todos los aspectos concretos, sea el código empleado propietario o no. En la medida en que, como se ha reiterado, estamos hablando de un *código* que es a la postre *código jurídico*, toda esta labor ha de ser completamente controlada y definida

por la Administración pública, que es además la responsable a todos los efectos de la misma. Por estas razones, asumir la identidad reglamentaria, desde esta perspectiva funcional, de los algoritmos, puede servir para avanzar en la convicción de que hemos de establecer un marco normativo que garantice estos principios. Por ejemplo, en materia de transparencia y auditabilidad, donde estamos aún muy lejos de haber logrado una solución satisfactoria.

III. LAS SOLUCIONES

Frente a esta situación, el Derecho español, al igual que el resto de sistemas jurídicos de nuestro entorno, viene ensayando diversas soluciones jurídicas que, en la línea de garantizar una transparencia algorítmica de mínimos, a fin de permitir cierta auditabilidad de los programas empleados para ayudar a la realización de los quehaceres administrativos, son todavía manifiestamente insuficientes y deberán ser sensiblemente profundizadas. Todo sistema eficaz de control y supervisión de los instrumentos algorítmicos requiere, por lo demás, no sólo de ciertos niveles de transparencia respecto de su código fuente sino también de una explicabilidad suficiente como para que éstos sean comprensibles (Cerrillo i Martínez). Esta explicación o justificación de las decisiones públicas es relevante incluso si entendemos que los algoritmos no son funcionalmente normas reglamentarias sino actuaciones administrativas, ya que éstas también han de estar correctamente motivadas (art. 9.3 CE y art. 35 Ley 39/2015, de Procedimiento Administrativo Común -LPAC-) y para ello un mínimo de explicabilidad del algoritmo también será exigible. A partir de esta constatación repasamos rápidamente las soluciones ensayadas hasta la fecha y las que el Derecho europeo y español parecen ir tratando de decantar para un futuro próximo.

III.1. El marco normativo actual: exigencias de explicabilidad de mínimos, perspectivas de profundización y mejora y el recurso a las garantías en materia de protección de datos como complemento

Como se ha señalado, incluso desde paradigmas que entienden de una manera menos exigente el papel jurídico que realizan los programas y algoritmos que ayudan a la toma de decisiones administrativas o las conforman decisivamente

es necesario garantizar su justificación aceptable dentro del marco normativo actual.

Algo así es lo que trata de garantizar a día de hoy nuestro ordenamiento jurídico, donde no se considera a los algoritmos o programas que toman decisiones administrativas (ya sea de forma completa, ya condicionando de tal forma la decisión administrativa final que su efecto sea funcionalmente equivalente) como normas reglamentarias sino actuaciones muy singularizadas que requieren de una explicabilidad suficiente como para pasar el señalado filtro en términos de motivación. Se trata de una aproximación que, aunque insuficiente, es, por lo demás, relativamente común en Europa y en el Derecho comparado y acorde con los instrumentos internacionales existentes en la materia. Por ejemplo, la recomendación de la UNESCO de 2021 sobre la Ética de la Inteligencia Artificial (considerandos 37 a 41) se concentra, de nuevo, en una detallada exigencia de explicabilidad de la actuación algorítmica como alternativa a la completa transparencia de los algoritmos empleados para la adopción de decisiones públicas.

El Derecho español vigente, en el art. 41 de la Ley 40/2015, de Régimen Jurídico del Sector Público (LRJSP), se decanta por esta solución y lo hace, además, con notable contención a la hora de establecer controles y cautelas respecto del uso de algoritmos. Tras definir en su primer apartado lo que entiende que es una decisión administrativa automatizada (de forma, además, restrictiva, pues sólo lo será la que se realice “íntegramente a través de medios electrónicos”, de modo que la regulación vigente no se aplicará a los casos en que esta intervención sea auxiliar, complementaria o previa a la decisión final, por mucho que de facto sea determinante), el art. 41.2 LRJSP establece que en estos casos, únicamente, se deberá establecer previamente el órgano u órganos competentes “para la definición de las especificidades, programación, mantenimiento, supervisión y control de calidad y, en su caso, auditoría del sistema de información y de su código fuente”. El precepto también establece que ha de ser indicado el órgano considerado responsable a efectos de la impugnación de la actuación. Como puede verse, estamos ante unas exigencias completamente de mínimos respecto de lo que son las exigencias de transparencia y posibilidades reales de auditabilidad efectiva derivadas de ella. Por ello, ante una regulación que ha sido objeto de crítica, tanto por sus manifiestas insuficiencias como por la paradoja de que normas como la

Ley 12/2021, que regula las exigencias de explicabilidad de las empresas (plataformas digitales) que emplean a trabajadores estableciendo sistemas de control algorítmico tenga más obligaciones de explicabilidad que los propios poderes públicos (el nuevo art. 64.4 d) del Estatuto de los Trabajadores, al menos, sí obliga “a ser informado por la empresa de los parámetros, reglas e instrucciones en los que se basan los algoritmos o sistemas de inteligencia artificial que afectan a la toma de decisiones que pueden incidir en las condiciones de trabajo, el acceso y mantenimiento del empleo, incluida la elaboración de perfiles”, lo que es una exigencia bastante más concreta que la que tienen nuestras Administraciones Públicas, en un ejemplo de libro de la tendencia de nuestro ordenamiento jurídico, en ocasiones, a recurrir a la “*Drittwirkung* inversa” cuando a los poderes públicos ciertos controles no les gustan en exceso).

Esta situación, tan manifiestamente criticable, ha dado lugar al menos a dos reacciones para tratar de paliar el problema, aunque en ambos casos sin abandonar el paradigma de no exigencia de la completa transparencia y limitarse a reglas, aunque al menos algo más concretas, de explicabilidad.

La primera de ellas ha sido la aprobación, en 2021, de una Carta de Derechos Digitales por parte del Gobierno de España, aunque se trate de un documento sin contenido jurídico obligatorio y no excesivamente novedoso. La Carta, al menos, sí recopila muchas de las exigencias ya vigentes en nuestro Derecho y trata, además de profundizar en ellas. De este modo puede considerarse que, quizás, aporta pautas interpretativas que pueden ayudar a integrar una aplicación del art. 41.2 LRJSP algo más exigente. Si no, al menos, empieza a marcar una línea de mayor rigor, si bien de momento sólo *de lege ferenda*, que permite avizorar por dónde irán las reformas futuras en la materia. Así, la Carta establece, en su artículo XVIII, los derechos digitales de la ciudadanía en sus relaciones con las Administraciones Públicas, y en su apartado segundo se indica que “(e)l principio de transparencia y de reutilización de datos de las Administraciones públicas guiará la actuación de la Administración digital, de conformidad con la normativa sectorial. En particular, se garantizará el derecho de acceso a la información pública, se promoverá la publicidad activa y la rendición de cuentas y se velará por la portabilidad de los datos y la interoperabilidad de los formatos, sistemas y aplicaciones, en los términos que prevea el ordenamiento jurídico vigente”. Parece claro que el precepto

aspira a que en el futuro se normativice un derecho de la ciudadanía a esa publicidad activa, al menos en algunos casos, y a cierta transparencia, aunque ésta quede completamente por definir y a expensas de una ulterior concreción administrativa.

Algo más incisivo es el apartado sexto de ese mismo artículo XVIII, que establece que se ha de tratar de promover que la ciudadanía vaya teniendo reconocidos los derechos a que las decisiones administrativas estén guiadas por principios de ética pública en materia de IA, a obtener una motivación comprensible en lenguaje natural de las decisiones en entornos digitales, especialmente cuando éstas se apoyen en el criterio propuesto por un sistema automatizado o inteligente (como puede verse, se sigue en la lógica de la motivación de la decisión) y, muy especialmente, dice la propuesta, a la “transparencia sobre el uso de instrumentos de inteligencia artificial y sobre su funcionamiento y alcance en cada procedimiento concreto y, en particular, acerca de los datos utilizados, su margen de error, su ámbito de aplicación y su carácter decisorio o no decisorio”, lo que ya concreta en mayor medida. Asimismo, el precepto remite a la ley la decisión sobre las condiciones de transparencia y acceso, en su caso, al código fuente, con la finalidad esencial de garantizar que no se produzcan resultados discriminatorios que, como hemos mencionado, es uno de los riesgos que más atención han generado y que se han de conjurar. También se prevé, en principio, una reserva de las decisiones discrecionales a personas, en lo que es una interesante aplicación práctica de lo que Ponce Solé (2019) ha llamado “reserva de humanidad”. Como puede verse, la Carta de Derechos Digitales, en este punto, sin abandonar la lógica de primar la explicabilidad sobre la transparencia completa, al menos, concreta la primera algo más de lo que lo hace el marco regulatorio vigente (por lo que sería una muy buena noticia que estas previsiones fueran concretándose en mandatos normativos) y, respecto de la segunda, acepta que al menos en ciertos ámbitos y para ciertas cuestiones ésta ha de ser imprescindible para una completa y perfecta auditabilidad de los algoritmos que puedan suponer riesgos de discriminación. Estas previsiones, además, se completan con las del artículo XXV de la Carta, que lista una serie de derechos frente a la inteligencia artificial (no discriminación, condiciones de transparencia y buena gobernanza, accesibilidad, usabilidad y fiabilidad), haciendo hincapié, de nuevo, en que las personas tienen reconocido el derecho a solicitar una supervisión e intervención humana y a impugnar en todo caso las decisiones

automatizadas tomadas por sistemas de inteligencia artificial que produzcan efectos en su esfera personal y patrimonial.

Como se ha comentado, un segundo efecto interesante de esta Carta es que habrá de irse filtrando poco a poco a nuestro ordenamiento positivo. Y, de algún modo, aunque aún con timidez, es algo que estamos viendo ya. Así, la reciente Ley 15/2022, integral para la igualdad de trato y la no discriminación, establece en su art. 23 mecanismos para garantizar que no pueda haber discriminaciones derivadas del empleo de IA o de mecanismos automatizados de toma de decisiones, imponiendo a los poderes públicos la obligación de actuar de acuerdo con los principios de obligación de minimizar sesgos, transparencia y rendición de cuentas, así como la necesidad de proceder a evaluaciones de los resultados de aplicar estos sistemas en cuanto a sus posibles efectos discriminatorios. A estos efectos, se prevé tanto la creación de un sello de calidad de los algoritmos como la obligación de seguir las recomendaciones de la UE en materia de ética algorítmica. Pero, más allá del terreno de los principios, sobre todo, el art. 23.2 impone a las Administraciones Públicas la priorización de la transparencia en el diseño y la implementación de los algoritmos que toman decisiones, así como para explicar e interpretar las decisiones adoptadas por los mismos. Se trata de una cierta obligación jurídica de fines más que de resultado, pero obligación al fin y al cabo que, aunque no llega a ser tan exigente quizás como la de empleadoras y empleadores con respecto a las personas que trabajan en sus empresas que analizábamos antes, lo que no deja de ser paradójico, va en el buen camino. Es llamativo, eso sí, y criticable, que no se haya incorporado al art. 41.2 LRJSP.

Estas manifiestas insuficiencias respecto de las obligaciones de transparencia y explicabilidad de los algoritmos empleados por el sector público han tratado en ocasiones de paliarse aplicando las medidas que respecto de la utilización de sistemas de inteligencia artificial que afectan a la ciudadanía prevé el Reglamento General en materia de Protección de Datos (RGPD), norma europea de obligado cumplimiento que además ha sido continuada con la Ley Orgánica 3/2018, de Protección de Datos Personales y Garantía de los Derechos Digitales, de contenidos equivalentes. Ello no obstante, y a pesar de que el considerando 71 RGPD sí que explicita la necesidad de estas exigencias de transparencia para la auditabilidad de los sistemas algorítmicos empleados para la toma de decisiones, los efectos de aplicar normas en materia

de protección de datos a la actuación algorítmica de los poderes públicos son insuficientes, al menos, por dos motivos: el primero de ellos porque son normas pensadas más para el tráfico privado que la acción pública (de nuevo, estaríamos tratando de compensar fallos en la regulación de la actuación de nuestros poderes públicos intentando que al menos se sometan a los mismos controles que son propios del ámbito privado, cuando las garantías para la acción administrativa debieran ser diferentes y más intensas) y, en segundo término, porque además la regulación del art. 22 RGPD, como no puede ser de otra manera, establece como excepción a su regla, propia del ámbito de la protección de datos, que da a la ciudadanía derecho a no ser objeto de una decisión automatizada, precisamente el caso de que una norma jurídica autorice estas actuaciones para salvaguardar derechos o libertades o intereses públicos, que es precisamente lo que hacen preceptos como el art. 41.2 LRJSP. De modo que por mucho que de una interpretación conjunta del art. 22 RGPD con el art. 15 RGPD y sus exigencias de calidad en el tratamiento de los datos, combinadas con los contenidos del referido considerando 71, se puede tratar de deducir un sistema algo más ambicioso y con concretas obligaciones de explicabilidad que van más allá de lo que el Derecho vigente exige a nuestras Administraciones Públicas (Palma Ortigosa, 2019; Soriano Arnanz, 2021c), lo cierto es que a la postre son las normas de Derecho público las que van a determinar el régimen jurídico aplicable a los tratamientos de esta índole que realizan nuestras Administraciones Públicas y son éstas las que hemos de mejorar, introduciendo mayores exigencias. Esta mejora es ciertamente urgente, pues el marco jurídico vigente para la transparencia y auditabilidad de las actuaciones públicas algorítmicas o condicionadas por el empleo de inteligencia artificial, tal y como viene de ser resumido, es manifiesto que es ciertamente insuficiente, tanto desde una perspectiva ambiciosa como la anteriormente expuesta que los equiparara a normas reglamentarias como, incluso, desde postulados que parten de una visión mucho más restrictiva basada en las exigencias de auditabilidad y motivación, en la que se fundamenta nuestro actual modelo de control.

III.2. Propuestas para la mejora del control

Frente a esta situación, manifiestamente insuficiente, estamos asistiendo a cierta ebullición de propuestas alternativas para un mejor control, que normalmente no llegan al extremo de trasladar las exigencias de publicidad y

transparencia de las normas reglamentarias a los algoritmos y programas empleados por los poderes públicos para tomar decisiones pero que, al menos, sí pueden ayudar a la mejora de la actual situación. De manera necesariamente sintética, los referenciamos a continuación, para concluir este trabajo con las posibilidades de desarrollo futuro más probable que se avizoran en el horizonte jurídico de la IA empleada por los poderes públicos.

1. Prohibición completa o selectiva del empleo de herramientas algorítmicas para la toma de decisiones, al menos, en ciertos casos. Esta es la solución alemana para decisiones discrecionales actualmente vigente, la propuesta contenida en la Carta de Derechos Digitales del Gobierno de España de 2021 para estos mismos supuestos de ejercicio de potestades no regladas (aunque aún sin reflejo en el Derecho positivo español y no de modo tan taxativo, pues la Carta abre la posibilidad de admitir excepciones) y, sobre todo, la base de documentos como el producido por Naciones Unidas en 2021 sobre la privacidad en la era digital y, de modo mucho más interesante por las consecuencias jurídicas que está llamado a tener, la propuesta de la Unión Europea de 2021 de Reglamento para la Regulación de la Inteligencia Artificial que, como es sabido, prevé directamente ámbitos en los que entiende que por su especial peligrosidad el empleo de herramientas algorítmicas habría de estar directamente prohibido en todo caso (para el ámbito público, la propuesta prohibiría todo uso de IA en cuanto a las funciones públicas de control de las poblaciones que pudiera comportar clasificaciones en categorías de ciudadanos a efectos de orden público).

Como puede verse, esta aproximación, que como ya hemos señalado nos retrotrae a las ideas defendidas por Ponce Solé de que ha de haber en el ámbito de lo público espacios sometidos a una “reserva de humanidad”, en los que la interacción deba ser humana y, por supuesto, la decisión final pesada y ponderada por inteligencias humanas, requiere de una previa decisión, de un previo acuerdo, sobre qué ámbitos estarán excluidos, lo que obliga a un proceso de deliberación democrática que analice pros y contras, riesgos y promesas de incrementos de eficacia, para delimitar las concretas fronteras de los ámbitos que deseamos que queden fuera de las capacidades de juicio de la IA. Es ésta muy probablemente la principal dificultad que presentan estas propuestas, por cuanto no es dudoso que en el futuro la evolución tecnológica prometerá ganancias de eficiencia también en estos ámbitos, y muy particularmente en muchos donde hasta la

fecha hay ejercicios de potestades discrecionales, obligando a una constante puesta en cuestión de estos límites. Por el contrario, estas aproximaciones tienen la ventaja indudable de que, una vez tomada la decisión, la puesta en marcha de la misma, su aplicación, es muy sencilla.

2. Mejoras de la auditabilidad por medio de sistemas avanzados de ingeniería inversa. Los sistemas de ingeniería inversa para lograr una mejor auditabilidad de los sistemas de IA están a la orden del día y son de probada utilidad para detectar problemas y sesgos. La cuestión es que, en ocasiones, no se facilita su empleo, que no requiere de la transparencia del código fuente pero sí de poder utilizar la herramienta con pocas limitaciones y poder probar con ella alternativas y contrafácticos. Hay, por ello, propuestas muy interesantes, como la de Sandra Wachter, de compensar la no transparencia del código fuente con normas jurídicas que permitan al menos a la ciudadanía o a asociaciones hacer este uso de los algoritmos para detectar problemas (Wachter *et al.*, 2018). Por poner un ejemplo pedestre pero ilustrativo, no hace falta acceder al código fuente del programa PADRE de la Agencia Tributaria para detectar problemas en el mismo, por cuanto el hecho de que toda la ciudadanía use intensivamente el programa hace que cualquier incoherencia del mismo, tanto en su funcionamiento interno como en posibles disfunciones a la hora de trasladar las previsiones normativas sobre tributación de modo que no respete la programación lo que establecen las leyes que regulan el impuesto van a ser fácilmente detectadas por la propia dinámica de uso de los y las ciudadanas, que lo tienen a su disposición para probar cómo reacciona el programa ante la introducción de datos en un sentido u otro.

Poder tener las herramientas algorítmicas que emplea la Administración pública a disposición de la ciudadanía para hacer pruebas con ellas introduciendo contrafácticos podría ser, por ello, una medida muy interesante para paliar los problemas que se deducen de la no existencia de una completa transparencia y publicación del código fuente de la herramienta. Se trata de una solución con no pocas ventajas, pues es sencilla de poner en marcha, preserva los posibles derechos propietarios sobre las herramientas de IA empleadas, requiere de pocos o ningún ajuste normativo y, además, puede ser empleada para cualquier herramienta, implicando a la sociedad civil y a colectivos ya demostradamente comprometidos con estas tareas. Sus únicos inconvenientes son que para poder llevar a cabo estos controles las personas

usuarias requieren de una capacidad técnica especializada (lo que hace que las posibilidades de control efectivo recaigan sólo sobre parte de la sociedad, que además ha de organizarse para ello, y no sobre un control ciudadano poroso más general), que siempre cabe la posibilidad de que ciertos problemas tarden en ser detectados y, sobre todo, que aquellos sesgos más complejos de detectar puedan permanecer ocultos porque la capacidad de cualquier sistema de ingeniería inversa en términos de auditabilidad no deja de ser siempre subóptima respecto de la completa transparencia.

3. Profundización en el modelo de explicabilidad reforzada. Una tercera alternativa, en la línea que ya hemos referido respecto del Derecho español y su evolución, es la mejora, en diferentes grados de ambición, del marco legal vigente basado en la auditabilidad a partir de la explicabilidad. Así, la ya referida Recomendación de la UNESCO de 2021 sobre la Ética de la Inteligencia Artificial y, sobre todo, la propuesta de la UE de Reglamento de Inteligencia Artificial de 2021 van en esta línea, incrementando las exigencias de información y transparencia mínimas para los usuarios. Es el caso del art. 10 de la propuesta de Reglamento europeo, que detalla obligaciones sobre cómo han de diseñarse y funcionar los algoritmos, o de su art. 13, que lista unas exigencias mínimas de información que han de estar en todo caso siempre disponibles para los usuarios (o ciudadanos) respecto de los sistemas de inteligencia artificial que se empleen con ellos, dependiendo del tipo de sistema, y que incluiría requerimientos técnicos diversos, resultados de evaluaciones y una detallada explicación de la lógica del sistema.

Una de las herramientas propuestas con el objetivo de mitigar algunos de los riesgos generados por el creciente uso de sistemas automatizados es el desarrollo y establecimiento de bases de datos o registros públicos en los que se identifiquen y se proporcione información acerca de diferentes sistemas algorítmicos (Soriano Arnanz, 2021a, ha propuesto la creación en España de estos registros a partir de ejemplos como los de Ámsterdam o Helsinki). En este sentido, el artículo 60 de la propuesta, en estos momentos en proceso de debate a nivel europeo, de Reglamento de Inteligencia Artificial de la Unión Europea prevé el establecimiento de una Base de datos de la UE para sistemas de Inteligencia Artificial de alto riesgo. En la misma línea encontramos, además, determinados documentos de *soft-law* que se han ido aprobando en los últimos años, como las ya referidas Carta de Derechos Digitales del Gobierno

de España presentada en 2020 o la Recomendación de la UNESCO para el empleo de la Inteligencia Artificial de 2021. Ambos instrumentos, de forma coherente con la propuesta de reglamento para el control y encuadramiento jurídico de los usos de la inteligencia artificial (también por parte de los poderes públicos) inciden mucho en la necesidad de establecer estrictas obligaciones de explicabilidad y transparencia, en la línea de las medidas paliativas de mínimos que, de acuerdo con algunas propuestas *de lege ferenda* (Boix Palop, 2021; Soriano Arnanz, 2021d) que han aparecido en relación con la necesidad de adopción de ciertas medidas de control, ya podemos ir trazando, si bien aún sin todo el grado de desarrollo y concreción, al menos por el momento, deseable. Por esta razón, la puesta en práctica de estas actuaciones, a partir de los primeros desarrollos prácticos que pueda llevar a cabo la Administración Pública, concretando estos modelos, revestirá de una gran importancia, al marcar en un plano aplicativo la línea a seguir y suponer una verdadera prueba de las posibilidades y límites de estos mecanismos de control.

Un ejemplo de iniciativas en esta línea, de hecho la primera en España, es el artículo 16.1.l) de la Ley 1/2022, de 13 de abril, de Transparencia y Buen Gobierno de la Comunitat Valenciana, que establece que las Administraciones públicas valencianas deberán publicar “la relación de sistemas algorítmicos o de inteligencia artificial que tengan impacto en los procedimientos administrativos o la prestación de los servicios públicos con la descripción de manera comprensible de su diseño y funcionamiento, el nivel de riesgo que implican y el punto de contacto al que poder dirigirse en cada caso, de acuerdo con los principios de transparencia y explicabilidad”. Esta norma, además, certeramente, y a diferencia del art. 41.2 LRJSP, obliga a incluir en el registro a todos los algoritmos empleados que directa o indirectamente afectan o condicionan la decisión administrativa final, sin limitarse sólo a las decisiones exclusivamente algorítmicas, y atendiendo a la efectiva función de los usos de la IA y su influencia real, lo que es la aproximación correcta. Con todas las posibilidades de mejora, se trata de un primer paso muy relevante, que es de esperar que sea imitado, y que debería ser profundizado con una ampliación de la exigibilidad jurídica de los contenidos mínimos que ha de contener el registro.

4. Desarrollo de modelos de auditoría externa, ya sea de tipo privado, ya a cargo de instituciones públicas. Una cuarta línea, que está siendo explorada especialmente en

los Estados Unidos, es la que a partir de mecanismos de autorregulación, y con la idea de preservar en todo caso los sistemas propietarios del escrutinio público generalizado que se derivaría de obligarles a la publicación del código fuente, trata de establecer como alternativa a la transparencia la auditoría realizada por entes u organismos acreditados para esta labor, externos a los poderes públicos, a los que sí se daría acceso al código de forma completa. Obviamente, este sistema plantea algunos problemas, tanto de posible colusión como de externalización de una labor que debiera ser bien pública, bien ciudadana. Por esta razón, en el contexto europeo suele plantearse como alternativa que esta actividad de control, incluyendo el acceso al código fuente de las herramientas de IA que emplee el poder público, incluso en el caso de que sean *software* propietario con protección de su propiedad intelectual, sea realizada por organismos públicos especializados, bien de nueva creación, bien ya existentes a los que se atribuyan estas competencias adicionales (Hoffmann-Riem [2018] ha divulgado mucho este tipo de propuestas, que en España han tendido a recogerse defendiendo la conveniencia de radicar estas funciones en las autoridades de protección de datos).

Con todo, este despliegue ambicioso de instituciones públicas de control no ha llegado aún a fructificar con toda su intensidad y, por ejemplo, la propuesta europea de Reglamento de Inteligencia Artificial de 2021 ya referenciada, aun estableciendo una serie de autoridades administrativas a las que debería informarse del empleo de IA y de todas las cuestiones técnicas referidas a estos sistemas, no llega al extremo de concebir el código fuente, siempre, por defecto y en todo caso, entre la información que haya de notificarse a estas autoridades públicas para su supervisión y control sin admitir excepciones (art. 30 de la Propuesta). En cualquier caso, y como alternativa a una total transparencia de los algoritmos o sistemas de IA empleados por los poderes públicos, esta línea de trabajo puede resultar muy productiva, al combinar la preservación de los derechos e intereses de los desarrolladores y propietarios de los sistemas con mecanismos de control público que podrían llegar a ser suficientemente exigentes si se desplegaran de forma ambiciosa (Cotino Hueso, 2022). Cuestión distinta es que no deja de ser, de nuevo, un sistema de control subóptimo comparado con la total publicación del código. Piénsese, por ejemplo, que estos organismos, si bien cuando controlaran el código fuente de las herramientas de IA que pudieran emplear empresas o sujetos privados en sus relaciones con otros ciudadanos lo harían desde una posición de supervisión y control

externo, cuando hicieran lo propio con algoritmos o programas empleados por el poder público, aun desde una posición de independencia y autonomía garantizadas, no dejarían de estar procediendo a este control como parte del sector público y del entramado de poder institucional.

Como puede verse, y aunque la total transparencia no sea aún la solución mayoritariamente defendida, la conciencia sobre las enormes y peligrosas deficiencias del estado regulatorio actual en punto a los controles sobre la IA empleada por las Administraciones Públicas es ya compartida de forma general, razón por la cual asistimos a este momento de ebullición con la aparición de muy diversas propuestas, que abren distintas posibilidades futuras. No es posible aún saber en qué cristalizarán las diferentes concepciones que hemos comentado, aunque parece fácil aventurar que el modelo al que vamos a hibridar de algún modo. Así, es posible concebir que, al menos a corto plazo, el Derecho europeo impulsará en la línea reseñada la profundización en las exigencias de explicabilidad (esto es, una transparencia limitada, que normalmente no incluirá la publicidad por defecto generalizada del código fuente), probablemente complementada con un incremento de las exigencias respecto del contenido mínimo de la explicabilidad que ha de proporcionarse, al estilo de lo que ya vemos en algunas normas o prácticas más ambiciosas, y a la que podrían incorporarse reglas de obligado cumplimiento para facilitar la ingeniería inversa, el complemento de abrir posibilidades para la auditoría de códigos fuente a cargo de entes privados, así como, por último, un paulatinamente creciente protagonismo de las futuras agencias públicas de notificación en funciones de supervisión y control. Todo ello unido a lo que debiera ser la tendencia de futuro y que, esperemos, acabe por imponerse, que es exigir cada vez en más y más ámbitos y respecto de más y más ejemplos de uso de la IA por parte del poder público, la completa publicidad de su código fuente. En todo caso, para llegar a este punto es evidente, a la luz del panorama descrito, que queda aún mucho camino por recorrer, visto lo visto y analizadas las muy escasas exigencias con las que en la actualidad enmarca nuestro Derecho positivo la actuación de nuestros poderes públicos cuando utilizan programas o algoritmos que contengan uso de IA para la adopción de decisiones que afectan a la ciudadanía. Una situación que hay que ir revirtiendo, quizás poco a poco y con prudencia, velando en todo caso por la corrección jurídica y tratando de no perjudicar de manera muy gravosa el despliegue de herramientas que pueden mejorar la vida de los ciudadanos,

pero sin demora. Porque la situación actual de falta de transparencia no es, sencillamente, admisible en un Estado de Derecho avanzado.

IV. BIBLIOGRAFÍA

- BARONA VILAR, S., (2022): “La seductora algoritmización de la justicia. Hacia una justicia poshumanista (Justicia+) ¿utópica o distópica?”, *El Cronista del Estado Social y Democrático de Derecho*, núm. 100, pp. 36-47.
- 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”, *Teoría y Método: Revista de Derecho Público*, núm. 1, pp. 223-269.
- CAMPUZANO FERNÁNDEZ, S. y VEGA VICENTE, P., (2021): “Riesgos y oportunidades de la inteligencia artificial desde la perspectiva de la competencia”, *Boletín Económico ICE*, núm. 3137, pp. 43-55.
- 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.
- COTINO HUESO, L., (2020): "SYRI, ¿A quién sanciono? Garantías frente al uso de la inteligencia artificial y decisiones automatizadas en el sector público y la sentencia holandesa de febrero de 2020", *La Ley Privacidad*, núm. 4.
- (2022): “Transparencia y explicabilidad de la inteligencia artificial y “compañía” (comunicación, interpretabilidad, inteligibilidad, audita-bilidad, testabilidad, comprobabilidad, simulabilidad...). Para qué, para quién y cuánta”, en: Cotino Hueso, L. y Castellanos, J. (coords.), *Transparencia y explicabilidad de la inteligencia artificial*, València, Tirant lo Blanch.
- (2022): “Sistemas de inteligencia artificial con reconocimiento facial y datos biométricos: mejor regular bien que prohibir mal”, *El Cronista del Estado Social y Democrático de Derecho*, núm. 100, pp. 68-79.

- DE LA CUEVA, J., (2019): “¿Quién vigila al algoritmo?”, *El Notario del siglo XXI*, núm. 87.
- FERNÁNDEZ TORRES, J. R., (1998): *Jurisdicción administrativa revisora y tutela judicial efectiva*, Madrid, Civitas.
- GARCÍA DE ENTERRÍA, E., (1962): “La lucha contra las inmunidades del poder en el Derecho administrativo”, *Revista de Administración Pública*, núm. 38, pp. 159-208.
- GOERLICH PESET, J. M., (2022): *Algoritmos en las relaciones laborales: control empresarial y control de las empresas*, “Impacto de la digitalización en los estudios de Derecho, Ciencia Política y Criminología”, 11 de mayo, Valencia.
- HOFFMANN-RIEM, W., (2018): *Big Data. Regulative Herausforderungen*, Baden-Baden: Nomos, 2018; traducción al castellano: *Big data. Desafíos también para el Derecho*, Madrid, Civitas.
- HUERGO LORA, A., (2021): *La regulación de los algoritmos*, Cizur Menor (Navarra), Aranzadi Thomson-Reuters.
- MARTÍN DELGADO, I., (2009): “Naturaleza, concepto y régimen jurídico de la actuación administrativa automatizada”, *Revista de Administración Pública*, núm. 180, 2009, pp. 353-386.
- MARTÍNEZ GARAY, L., (2018): “Peligrosidad, algoritmos y due process: El caso State vs. Loomis”, *Revista de Derecho Penal y Criminología*, núm. 20, pp. 485-502.
- MARTÍNEZ GARAY, L. y MONTES SUAY, F., (2018): “El uso de valoraciones del riesgo de violencia en Derecho Penal: algunas cautelas necesarias”, *Indret*, núm. 2.
- MARTÍNEZ GARAY, L. y GARCÍA ORTIZ, A., (2022): “Paradojas de los algoritmos predictivos utilizados en el sistema de justicia penal”, *El Cronista del Estado Social y Democrático de Derecho*, núm. 100, pp. 160-173.

- MESTRE DELGADO, J., (1998), *El derecho de acceso a archivos y registros administrativos (análisis del art. 105 b) de la Constitución*, Cizur Menor (Navarra), Aranzadi.
- (2018): “Artículo 105”, en: Rodríguez-Piñero Bravo-Ferrer, M. y Casas Baamonde, M. E. (Dirs.), *Comentarios a la Constitución Española*, Madrid, BOE-Wolters-Kluwer.
- O’NEIL, C., (2016): *Weapons of Math Destruction*, Londres, Allen Lane.
- OROFINO, A. G., (2020): “The Implementation of the Transparency Principle in the Development of Electronic Administration”, *European Review of Digital Administration & Law*, núm. 1-2, pp. 123-142.
- PALMA ORTIGOSA, A., (2019): “Decisiones automatizadas en el RGPD. El uso de algoritmos en el contexto de la protección de datos”, *Revista General de Derecho Administrativo*, núm. 50.
- PRESNO LINERA, M. A., (2022): “Derechos fundamentales e inteligencia artificial en el estado social, democrático y digital de derecho”, *El Cronista del Estado Social y Democrático de Derecho*, núm. 100, pp. 48-57.
- 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.
- RANCHORDÁS, S. y SCHUURMANS, Y., (2020): “Outsourcing the Welfare State: The Role of Private Actors in Welfare Fraud Investigations”, *European Journal of Comparative Law and Governance*, pp. 5-42.
- RODRÍGUEZ PEÑA, N. L., (2021): “Big data e inteligencia artificial: una aproximación a los desafíos éticos y jurídicos de su implementación en las administraciones tributarias”, *Ius et Scientia*, vol. 7, núm. 1, pp. 62-84.
- SÁNCHEZ GALÁN, L. y REGAL RAMOS, R., (2019): “Recaídas en incapacidad temporal: impacto de su regulación y control”, *Medicina y Seguridad del Trabajo*, vol. 65, núm. 256, pp. 217-232.

- SORIANO ARNANZ, A. (2021a): “Decisiones automatizadas: problemas y soluciones jurídicas. Más allá de la protección de datos”, *Revista de Derecho Público: teoría y método*, núm. 1.3, pp. 85-127.
- (2021b): Decisiones automatizadas y discriminación: aproximación y propuestas generales, *Revista General de Derecho Administrativo*, núm. 56.
- (2021c): *Data Protection for the prevention of algorithmic discrimination*, Cizur Menor (Navarra): Aranzadi Thomson Reuters.
- (2021d): La propuesta de Reglamento de Inteligencia Artificial de la Unión Europea y los sistemas de alto riesgo, *Revista General de Derecho de los Sectores Regulados*, núm. 8.
- (2022): "El uso de sistemas automatizados por los poderes públicos. El caso del sistema empleado por el servicio público de empleo austriaco", en: Cotino Hueso, Lorenzo y Todolí Signes, Adrián (Coords.), *Explotación y regulación del uso del big data e inteligencia artificial para los servicios públicos y la ciudad inteligente*, València: Tirant Lo Blanch, pp. 192-216.
- SORIANO ARNANZ, A. y SIMÓ SOLER, E., (2021): *Machine learning y Derecho: aprendiendo la (des)igualdad*, Barona Vilar, S. (Dir.), Valencia, Tirant lo Blanch, pp. 183-207.
- 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*, núm. 58, pp. 82-96.
- VESTRI, G., (2021): “La inteligencia artificial ante el desafío de la transparencia algorítmica: una aproximación desde la perspectiva jurídico-administrativa”, *Revista Aragonesa de Administración Pública*, núm. 56, pp. 368-398.
- WACHTER, S., MITTELSTADT, B. y RUSSELL, C. (2018): “Counterfactual explanations without opening the black box: automated decisions and the GDPR”, *Harvard Journal of Law & Technology*, vol. 31, núm. 2, pp. 841-887.

- ZARSKY, T., (2013): “Transparent Predictions”, *Illinois Law Review*, núm. 4, pp. 1503-1570.
- (2016): “The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making”, *Science, Technology & Human Values*, vol. 41, núm. I, pp. 118-132.