

UNA PANORÁMICA DE LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL DESDE LA PERSPECTIVA DEL DERECHO ADMINISTRATIVO^(*)(^{**})

JESÚS A. TAHIRÍ MORENO
Doctorando de Derecho Administrativo
Universidad de Zaragoza

SUMARIO: I. INTRODUCCIÓN AL CONCEPTO DE SISTEMA DE INTELIGENCIA ARTIFICIAL: 1. Conceptos fundamentales y taxonomía de los sistemas de inteligencia artificial. A) Concepto. B) Inteligencia artificial general o fuerte. C) Inteligencia artificial débil o estrecha. D) Inteligencia artificial simbólica y subsimbólica o conexionista. E) Los modelos fundacionales. F) Ultrafalsificaciones o deepfakes. 2. La relación género-especie entre sistemas automatizados y sistemas de inteligencia artificial.– II. APROXIMACIÓN AL APRENDIZAJE AUTOMÁTICO Y A LAS DIFERENTES TÉCNICAS EMPLEADAS EN EL DESARROLLO DE MODELOS PREDICTIVOS: 1. ¿Qué es el aprendizaje automático o machine learning? 2. Las diferentes técnicas de aprendizaje automático. A) Aprendizaje supervisado. B) Aprendizaje no supervisado. C) Aprendizaje por refuerzo. 3. La inteligencia artificial generativa. III. LOS RIESGOS Y OPORTUNIDADES DERIVADOS DEL USO DE SISTEMAS DE INTELIGENCIA ARTIFICIAL EN LA ADMINISTRACIÓN PÚBLICA: 1. Una exposición de los principales riesgos derivados del uso de sistemas de inteligencia artificial en la actuación administrativa. A) La discriminación algorítmica. B) El sesgo de automatización. C) La falta de transparencia de los algoritmos. 2. Las oportunidades y ventajas de su implementación en la Administración pública. A) Incremento de la eficacia y de la eficiencia en la prestación de los servicios públicos. B) La personalización de los servicios públicos.– IV. EJEMPLOS DE PROGRAMAS INFORMÁTICOS QUE EMPLEAN INTELIGENCIA ARTIFICIAL EN LA ACTIVIDAD ADMINISTRATIVA. 1. La aplicación informática del Cuerpo Nacional de Policía «Veripol». 2. El sistema «KENDOJ» del Centro de Documentación Judicial. 3. El asistente virtual «ISSA» del Instituto Nacional de la Seguridad Social. 4. La herramienta de valoración del riesgo de la administración penitenciaria catalana «RisCanvi».– V. LA REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL A NIVEL EUROPEO Y AUTONÓMICO: 1. Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos

(*) Trabajo recibido en esta REVISTA con fecha 10/10/2023 y evaluado favorablemente para su publicación el 09/01/2024.

(**) Trabajo realizado en el marco del Proyecto de investigación PID2021-127828NB-I00 «Innovación para una salud de vanguardia: compra pública, tecnología, sostenibilidad ambiental y factores socioeconómicos (IN-SALVAN)» y el Proyecto PID2022-136548NB-I00 «Los retos de la inteligencia artificial para el Estado social y democrático de Derecho», financiados por el Ministerio de Ciencia, Innovación y Universidades.

de la Unión. A) Sistemas prohibidos (de riesgo inaceptable). B) Sistemas de alto riesgo. C) Sistemas de riesgo limitado. 2. Decreto-ley 2/2023, de 8 de marzo, de medidas urgentes de impulso a la inteligencia artificial en Extremadura.– VI. CONSIDERACIONES FINALES.– VII. BIBLIOGRAFÍA.

RESUMEN: Este trabajo nace con la vocación de facilitar la comprensión de los conceptos fundamentales sobre la regulación de la inteligencia artificial a quienes se aproximan por primera vez a la materia. A pesar de que las publicaciones sobre inteligencia artificial en la esfera del derecho son numerosas, la mayoría, o bien están dirigidas a una audiencia que cuenta con conocimientos previos del sector; o bien, si se dirigen a un público profano, no abordan la cuestión desde la perspectiva jurídica. El artículo, en consecuencia, tiene un cariz introductorio y pretende aportar al lector una visión general de los conceptos y problemáticas elementales de esta tecnología en un lenguaje claro y didáctico.

Palabras clave: inteligencia artificial; derecho administrativo; derecho europeo; aprendizaje automático; sistemas de inteligencia artificial.

ABSTRACT: The purpose of this paper is to facilitate the comprehension of basic concepts of artificial intelligence regulation for those approaching the field for the first time. Notwithstanding there are numerous publications on artificial intelligence in the field of law, most of them are either addressed to a public that already has prior knowledge of the sector or, if they are intended to a ignorant audience, they do not deal with the subject from a legal perspective. Consequently, the paper is introductory in nature and aims to provide the reader with an overview of the fundamental concepts and problems of this technology in a straightforward and didactic language.

Key words: artificial intelligence; administrative law; european law; machine learning; artificial intelligence systems

I. INTRODUCCIÓN AL CONCEPTO DE SISTEMA DE INTELIGENCIA ARTIFICIAL

1. Conceptos fundamentales y taxonomía de los sistemas de inteligencia artificial

La delimitación del concepto de sistema de inteligencia artificial (en adelante, IA) tiene una significación jurídica de primer orden. Como acertadamente señala GAMERO CASADO, no toda actuación administrativa automatizada (1)

(1) Siguiendo el tenor literal del art. 41.1 de la Ley 40/2015 de la Ley de Régimen Jurídico del Sector Público (LRJSP) se entiende por actuación administrativa automatizada «cualquier acto o actuación realizada íntegramente a través de medios electrónicos por una Administración Pública en el marco de un procedimiento administrativo y en la que no haya intervenido de forma directa un empleado público».

implica el uso de un sistema de inteligencia artificial (2). En este sentido, debe tenerse en cuenta que el ámbito de aplicación del futuro Reglamento de Inteligencia Artificial (3) se circunscribe a los sistemas de IA en *stricto sensu*.

Actualmente en la doctrina iuspublicista no existe un criterio unánime y, en ausencia de una norma europea que aporte una definición común para todos los Estados miembro (4), nos encontramos con que cada autor califica un sistema según su leal saber y entender (5). Ante esta situación de inseguridad y con la finalidad de evitar lagunas, COTINO HUESO ha planteado un concepto inclusivo de IA aplicable al sector público que comprenda también a los sistemas automatizados (6).

Por otra parte, me parece interesante destacar que con el objetivo de simplificar y homogeneizar el lenguaje técnico relativo a la IA, la Asociación Española de Normalización (UNE) ha ratificado en el mes de agosto la norma técnica internacional «Tecnología de la información. Inteligencia artificial. Conceptos y terminología de inteligencia artificial (ISO/IEC 22989:2022)», aprobada por el Comité Europeo de Normalización (CEN) el 26 de junio de 2023 (7). Esta norma (8), a pesar de no ser jurídicamente vinculante, aporta una serie de conceptos con la finalidad de que todas las partes interesadas en la investigación, el desarrollo, la comercialización o la regulación (*stakeholders* para quien prefiera los anglicismos) de esta tecnología utilicen un lenguaje estandarizado (9).

(2) E. GAMERO CASADO (2023: 4).

(3) Propuesta de Reglamento del Parlamento Europeo y el Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión [COM(2021) 206 final].

(4) El pasado 8 de diciembre la Comisión, el Consejo y el Parlamento Europeo alcanzaron un acuerdo sobre el contenido del texto definitivo del Reglamento. En la fecha en que se escriben estas páginas aún no se ha redactado el documento. Una aproximación a las novedades que aporta el acuerdo provisional al proyecto de Reglamento inicial la encontrarán en el apartado 1 del epígrafe V del presente estudio.

(5) Esta situación se da principalmente en algunos sistemas utilizados por la Administración pública debido a la ausencia de información técnica.

(6) L. COTINO HUESO (2023: 213).

(7) El código español de la norma (UNE-EN ISO/IEC 22989:2023) ha sido publicado en el BOE a través de la Resolución de 4 de septiembre de 2023, de la Dirección General de Industria y de la Pequeña y Mediana Empresa, por la que se publica la relación de normas europeas que han sido ratificadas durante el mes de agosto de 2023 como normas españolas.

(8) El art. 8, apartado 3, de la Ley 21/1992 de Industria define una norma técnica como «La especificación técnica de aplicación repetitiva o continuada cuya observancia no es obligatoria, establecida con participación de todas las partes interesadas, que aprueba un Organismo reconocido, a nivel nacional o internacional, por su actividad normativa».

(9) Para profundizar en el fenómeno de la normalización industrial, véase V. ÁLVAREZ GARCÍA (2010): *Industria*, lustel, pp. 129-287.

A) Concepto

A escala internacional, la Organización para la Cooperación y el Desarrollo Económicos (OCDE) define un sistema de inteligencia artificial como «un sistema basado en una máquina que puede, para un conjunto determinado de objetivos definidos por humanos, hacer predicciones, recomendaciones o decisiones que influyan en entornos reales o virtuales. Los sistemas de IA están diseñados para operar con diferentes niveles de autonomía» (10). Esta definición ha sido acogida (con matices) por las instituciones europeas, tanto por la Comisión (en el proyecto de Reglamento de IA) como por el Parlamento Europeo (11) y el Consejo (12). También por los organismos internacionales y nacionales de normalización, quienes han concluido que un sistema de IA es un «Sistema de ingeniería que genera resultados como contenidos, previsiones, recomendaciones o decisiones para un conjunto determinado de objetivos definidos por el ser humano» (13) (la traducción al español es propia).

Otro concepto lo encontramos en la disposición 2, letra a), de la Recomendación sobre la Ética de la Inteligencia Artificial adoptada por la UNESCO. Para este organismo, los sistemas de IA «son tecnologías de procesamiento de la información que integran modelos y algoritmos que producen una capacidad para aprender y realizar tareas cognitivas, dando lugar a resultados como la predicción y la adopción de decisiones en entornos materiales y virtuales. Los sistemas de IA están diseñados para funcionar con diferentes grados de autonomía, mediante la modelización y representación del conocimiento y la explotación de datos y el cálculo de correlaciones» (14).

Ya en el plano del derecho interno, el Real Decreto 817/2023, de 8 de noviembre, que establece un entorno controlado de pruebas para el ensayo del cumplimiento de la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de

(10) Los Principios de la OCDE sobre Inteligencia Artificial pueden encontrarse en el siguiente enlace <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.

(11) Enmiendas aprobadas por el Parlamento Europeo el 14 de junio de 2023 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)).

(12) Puede accederse al documento que contiene la Orientación General del Consejo Europeo de 25 de noviembre de 2022 a través del siguiente enlace <https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/es/pdf>.

(13) Especificación 3.1.4. de la norma UNE-EN ISO/IEC 22989:2023.

(14) Acceso al texto completo en https://unesdoc.unesco.org/ark:/48223/pf0000380455_spa.

inteligencia artificial (15), define en su artículo 3, ap. 3 esta clase de *software* como un «sistema diseñado para funcionar con un cierto nivel de autonomía y que, basándose en datos de entradas proporcionadas por máquinas o por personas, infiere cómo lograr un conjunto de objetivos establecidos utilizando estrategias de aprendizaje automático o basadas en la lógica y el conocimiento, y genera información de salida, como contenidos (sistemas de inteligencia artificial generativos), predicciones, recomendaciones o decisiones, que influyan en los entornos con los que interactúa» (16).

A nivel autonómico, el artículo 4, letra c) del Decreto-ley 2/2023, de 8 de marzo, de medidas urgentes de impulso a la inteligencia artificial en Extremadura establece que un sistema de IA es «aquél *software*, que, empleando diferentes técnicas y estrategias para un conjunto determinado de objetivos definidos por seres humanos, puede generar información, contenidos, predicciones, recomendaciones o decisiones capaces de influir en los entornos con los que interactúa» (17).

En conclusión, de las diferentes definiciones expuestas pueden extraerse tres requisitos o presupuestos para calificar un sistema informático como sistema de IA. El primero, que pueda operar de forma autónoma (requisito de la autonomía). El segundo, el sistema tiene que ser capaz de generar información de salida que influya en el entorno en el que va a ser utilizado (requisito de la interactividad con el entorno). Por último, debe operar mediante técnicas de aprendizaje automático o estrategias basadas en la lógica y el conocimiento

(15) El Real Decreto tiene por objeto, a tenor de su art. 1, «establecer un entorno controlado de pruebas para ensayar el cumplimiento de ciertos requisitos por parte de algunos sistemas de inteligencia artificial que puedan suponer riesgos para la seguridad, la salud y los derechos fundamentales de las personas». Los entornos controlados de pruebas, también conocidos como *sandboxes* o cajas de arena (llamados así porque sus funciones se asimilan a las de los espacios de recreo para infantiles), permiten a los proveedores y usuarios de sistemas de IA de alto riesgo comprobar en una zona segura la operatividad de los requisitos establecidos en el Reglamento de IA, antes de que este entre en vigor. Ello se realizará mediante un proceso de autoevaluación del cumplimiento de los requisitos y la evaluación del plan de seguimiento posterior a la comercialización bajo la supervisión de la Secretaría de Estado de Digitalización e Inteligencia Artificial. Los resultados obtenidos en el entorno controlado se harán públicos y podrán ponerse a disposición de los organismos de normalización y la Comisión Europea. Además, de conformidad con la disposición adicional segunda, el órgano competente podrá desarrollar una *software* para una primera autoevaluación no vinculante partiendo de la información obtenida gracias al entorno controlado de pruebas.

(16) El concepto recogido en el Real Decreto 817/2023 es una fiel reproducción de la definición prevista en el art. 3.1 de la Orientación General del Consejo Europeo de 25 de noviembre de 2022.

(17) Al igual que el poder ejecutivo estatal, el legislador extremeño se ha limitado a reproducir, con matices, el concepto de sistema de IA del art. 3.1. del proyecto de Reglamento de IA.

que simulen o emulen la inteligencia humana (tecnología basada en IA en sentido estricto, ya opere con el paradigma simbólico o con el conexionista) (18).

B) Inteligencia artificial general o fuerte

La inteligencia artificial general (IAG) pretende alcanzar (e incluso superar) el nivel de capacidad cognitiva de los seres humanos. La finalidad de este modelo de IA es que las máquinas razonen, sientan y aprendan igual que nosotros. La IAG no ha sido aún desarrollada y su invención supondría igualar a humanos y máquinas. Los sistemas basados en IAG podrían sentir emociones y reflexionar sobre su propia existencia y el conocimiento que procesan. Hoy por hoy este tipo de IA es una utopía y en el estadio actual de la tecnología aún no se ha llegado a vislumbrar nada semejante.

C) Inteligencia artificial débil o estrecha

En esta categoría se integran todos los sistemas de IA que han sido desarrollados hasta la fecha. Utilizan lógica matemática, estadística o estrategias basadas en el conocimiento, están hiper especializados en la realización de tareas específicas y requieren grandes cantidades de datos para operar. A diferencia de la IAG, la IA débil no tiene la capacidad de comprender el conocimiento que procesa ni tiene consciencia sobre sí misma.

D) Inteligencia artificial simbólica y subsimbólica o conexionista

En el campo de la IA podemos diferenciar dos grandes enfoques o paradigmas: el simbólico (19) y el subsimbólico o conexionista (20).

(18) Una interpretación amplia del concepto de sistema de IA, a juicio del que escribe, sería contraproducente. El ámbito de aplicación de las normas que regulen este tipo de modelos debe constreñirse a aquellos que utilicen IA en su funcionamiento. La diferenciación jurídica entre sistemas automatizados en sentido estricto (que no emplean IA) y sistemas de IA es necesaria y conveniente, pues nada tienen que ver los riesgos potenciales que puede generar un sistema automatizado con los que puede llegar a provocar un modelo basado en IA. Son programas informáticos diferentes y la normativa reguladora debe prever tales diferencias. Sobre la cuestión, léase el apartado dos del presente capítulo.

(19) Siguiendo la definición dada por la especificación 3.1.34 de la norma UNE-EN ISO/IEC 22989:2023, la IA simbólica es una categoría de IA «basada en técnicas y modelos que manipulan símbolos y estructuras según reglas definidas explícitamente para obtener inferencias».

(20) La especificación 3.1.34 de la norma UNE-EN ISO/IEC 22989:2023 define el enfoque subsimbólico como un tipo de IA «basada en técnicas y modelos (3.1.23) que utilizan una codificación implícita de la información, que puede derivar de la experiencia o de los datos brutos».

En el enfoque clásico (IA simbólica), las máquinas llegan a un resultado concreto a través de la representación del conocimiento mediante símbolos. Para dar una información de salida (*output*), los sistemas basados en IA simbólica manipulan el conocimiento aplicando una serie de reglas lógicas predeterminadas por el desarrollador. Los sistemas expertos son el modelo de referencia de esta clase de IA (21). Técnicamente, el enfoque simbólico «consiste en codificar el conocimiento con símbolos y estructuras, utilizando principalmente la lógica para modelar los procesos de razonamiento. En este paradigma, la información es codificada explícitamente mediante una representación formal, cuya sintaxis es procesada por un ordenador y su semántica es comprendida por los humanos» (22).

En cuanto a la IA subsimbólica, también conocida como conexionista, parte de un enfoque en el que la máquina adquiere y procesa la información simulando el funcionamiento del cerebro humano a través de redes neuronales artificiales organizadas en capas ocultas. Dentro de esta categoría se incluyen los sistemas de IA que operan con técnicas de aprendizaje profundo (23) (*deep learning*). La diferencia entre la IA simbólica y subsimbólica reside, principalmente, en la forma en que representan el conocimiento ambos modelos. Así, en el enfoque simbólico las reglas han sido predeterminadas y la información se procesa mediante sistemas basados en la lógica y el conocimiento experto, mientras que en el modelo subsimbólico la información se extrae del análisis masivo de los datos y la identificación de patrones a través de correlaciones estadísticas.

E) Los modelos fundacionales

El concepto de modelo fundacional es reciente [apareció por primera vez en un informe de la Universidad de Stanford publicado en el año 2021 (24)] y está íntimamente relacionado con la IA generativa y la IA de uso o propósito general (25). En el texto que recoge las enmiendas del Parlamento Europeo

(21) Un sistema experto, es un «sistema de IA que almacena los conocimientos aportados por un experto humano en un dominio específico para inferir soluciones a problemas» (especificación 8.5.2 UNE-EN ISO/IEC 22989:2023).

(22) Especificación 5.9 de la norma UNE-EN ISO/IEC 22989:2023.

(23) Especificación 3.4.4 de la norma UNE-EN ISO/IEC 22989:2023.

(24) R. BOMMASANI et al. (2021: 3).

(25) El art. 3, apartado 1) ter de la Orientación General del Consejo de 25 de noviembre de 2022 define los sistemas de uso general como un «un sistema de IA que, con independencia de la manera en la que se introduzca en el mercado o se ponga en servicio, incluido el software de código abierto, ha sido concebido por el proveedor para desempeñar funciones de aplicación general, como el reconocimiento de imágenes y de voz, la generación de audio y vídeo, la detección de patrones, la respuesta a preguntas y la traducción, entre otras. Un sistema de IA de uso general puede utilizarse en una pluralidad de contextos e inte-

a la propuesta de Reglamento de IA de la Comisión se define un modelo fundacional como un «sistema de IA entrenado con un gran volumen de datos, diseñado para producir información de salida de carácter general y capaz de adaptarse a una amplia variedad de tareas diferentes» (26). Por su parte, el art. 3, apartado 6 del Real Decreto 817/2023, de 8 de noviembre lo define como «un modelo de inteligencia artificial entrenado en una gran cantidad de datos no etiquetados a escala (generalmente mediante aprendizaje autosupervisado y/o con recopilación automática de contenido y datos a través de internet mediante programas informáticos) que da como resultado un modelo que se puede adaptar a una amplia gama de tareas posteriores».

Estos sistemas se caracterizan por su flexibilidad y adaptabilidad para ser aplicados en diversos contextos. Un ejemplo paradigmático de esta nueva categoría de sistemas lo encontramos el modelo GPT (*Generative Pre-Trained Transformer* o transformador generativo preentrenado). Finalmente, conviene señalar que el acuerdo provisional alcanzado entre la Comisión, el Parlamento y el Consejo Europeo el 8 de diciembre de 2023 prevé normas específicas para este tipo de modelos, diferenciándose entre sistemas de IA de propósito general (GPAI por sus siglas en inglés), modelos fundacionales «ordinarios» (27) y modelos fundacionales de «alto impacto». Estos últimos no aparecían en ninguno de los textos previos y se distinguen, además de por la gran cantidad de datos empleados para su entrenamiento y su complejidad, por su capacidad para generar riesgos sistémicos a lo largo de toda la cadena de valor de la IA. De ahí que se haya decidido incluir en el futuro Reglamento un régimen más estricto para esta categoría de sistemas.

F) Ultrafalsificaciones o *deepfakes*

Define el Parlamento Europeo la ultrafalsificación como «un contenido de sonido, imagen o vídeo manipulado o sintético que puede inducir erróneamente a pensar que es auténtico o verídico, y que muestra representaciones de personas que parecen decir o hacer cosas que no han dicho ni hecho,

grarse en una pluralidad de otros sistemas de IA». Por su parte, el apartado 5 del art. 3 del Real Decreto 817/2023, de 8 de noviembre aporta una definición de los sistemas de IA de propósito general prácticamente idéntica a la prevista en la Orientación General del Consejo.

(26) Enmienda 168 del texto del Parlamento Europeo «Enmiendas aprobadas por el Parlamento Europeo el 14 de junio de 2023 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión (COM (2021)0206 – C9-0146/2021 /0106(COD))».

(27) En relación con los modelos fundacionales, la Comisión, el Parlamento y el Consejo han acordado regular obligaciones específicas en materia de transparencia antes de ser introducidos en el mercado.

producido utilizando técnicas de IA, incluido el aprendizaje automático y el aprendizaje profundo» (28).

De la mera lectura superficial del concepto ya se infiere que esta categoría de sistemas pueden vulnerar gravemente los derechos fundamentales y las libertades públicas de los ciudadanos (29). Por este motivo el proyecto de Reglamento de IA contempla medidas específicas de transparencia dirigidas a informar de la falsedad de los contenidos producidos (30). En particular, el art. 52, apartado 3 del proyecto de Reglamento impone a los usuarios de sistemas de IA que generen ultrafalsificaciones la obligación de comunicar que el contenido ha sido manipulado o generado de forma artificial (31).

2. La relación género-especie entre sistemas automatizados y sistemas de inteligencia artificial (32)

Como ya se adelantaba *ut supra*, la distinción entre sistemas autónomos o inteligentes (aquellos que emplean IA en su funcionamiento) y sistemas automatizados no es baladí para el derecho, pues en función de la naturaleza del programa informático serán de aplicación regímenes jurídicos diversos (33). En este sentido, debe partirse de una premisa: no todo sistema automatizado

(28) Enmienda 203 del texto del Parlamento Europeo «Enmiendas aprobadas por el Parlamento Europeo el 14 de junio de 2023 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión (COM (2021)0206 – C9-0146/2021 /0106(COD))».

(29) Un análisis claro y conciso sobre el impacto de la inteligencia artificial en los derechos fundamentales de los ciudadanos lo encontrará el lector en la monografía de M.A. PRESNO LINERA (2022): *Derechos fundamentales e inteligencia artificial*, Madrid, Marcial Pons.

(30) Recientemente un grupo de niñas menores de edad de Almendralejo (Badajoz) han sido acosadas y extorsionadas con imágenes generadas por una aplicación de IA cuya única finalidad es crear contenidos de «falsos desnudos». La noticia puede encontrarse en <https://www.xataka.com/privacidad/falsos-desnudos-menores-generados-ia-policia-investiga-almendralejo-primer-caso-masivo-espana>.

(31) Me parece interesante destacar que las grandes empresas tecnológicas (como Google, Meta y OpenAI, entre otras) ya están trabajando en el desarrollo de tecnologías cuya finalidad es etiquetar el contenido generado con IA mediante marcas de agua. Una muestra de ello la encontramos en la aplicación SynthID de Google. Esta herramienta permite añadir marcas de agua digitales invisibles a los píxeles de las imágenes generadas con IA.

(32) Para un análisis del régimen jurídico-administrativo de los sistemas automatizados desde la perspectiva del derecho comparado, véase la sección monográfica coordinada por O. MIR PUIGPELAT (2023): «La automatización y el uso de algoritmos e inteligencia artificial en Derecho administrativo comparado», *Revista General de Derecho Administrativo*, 63.

(33) Un estudio sobre la actuación administrativa automatizada y sus diferentes variantes lo encontrarán en A. CERRILLO MARTÍNEZ (2023): «Lección 21. Actuación automatizada, robotizada e inteligente», *Revista de Derecho Público: Teoría y método*, pp. 525-541.

es un sistema de IA. Ahora bien, todo sistema de IA sí queda subsumido en el concepto de sistema automatizado. La relación entre ambos conceptos es, pues, de género-especie (donde el sistema automatizado es el género y el sistema de IA es la especie).

Para saber si nos encontramos ante un sistema de IA lo ideal es conocer la información técnica del modelo. Las diferencias son fácilmente identificables (si se conocen los conceptos técnicos básicos) y atienden a las características funcionales del programa. Así, una aplicación que genere documentos de forma automática sin ningún nivel de autonomía (por ejemplo, cuando se expide la vida laboral desde la sede electrónica de la Seguridad Social o se presenta la declaración de la renta utilizando la aplicación de la Agencia Tributaria) no está empleando IA en su funcionamiento.

En la esfera de la Administración pública es importante que el personal a su servicio tenga una formación adecuada en inteligencia artificial que le permita diferenciar entre las distintas categorías de sistemas automatizados. Si la aplicación informática utilizada en la actuación administrativa no emplea técnicas de aprendizaje automático o estrategias basadas en la lógica y el conocimiento (IA en sentido estricto) no puede ser calificada como sistema de IA y, en consecuencia, no le será de aplicación la normativa que regule esta tecnología (34).

II. APROXIMACIÓN AL APRENDIZAJE AUTOMÁTICO Y A LAS DIFERENTES TÉCNICAS EMPLEADAS EN EL DESARROLLO DE MODELOS PREDICTIVOS

1. ¿Qué es el aprendizaje automático o *machine learning*?

El aprendizaje automático es una rama de la IA cuyo objeto consiste en desarrollar máquinas con la capacidad de aprender por sí mismas y mejorar su eficacia de forma autónoma a través de la experiencia. En palabras de GONZÁLEZ CABANES y DÍAZ DÍAZ, es el «área de estudio que busca emular en las máquinas la capacidad de aprender que tenemos humanos y seres vivos» (35).

(34) Un supuesto interesante que retrata a la perfección la problemática derivada de la confusión conceptual lo encontramos en los términos empleados por el legislador valenciano para delimitar los sistemas automatizados. Así, el art. 16.1.l) de la Ley 1/2022, de 13 de abril, de Transparencia y Buen Gobierno de la Comunitat Valenciana habla de «sistemas algorítmicos o de inteligencia artificial». La intención es evidente: aportar un concepto amplio que integre toda especie de sistema automatizado. No obstante, los sistemas de IA son también sistemas algorítmicos, por lo que su distinción es redundante. En mi opinión, hubiera sido más preciso desde el punto de vista técnico y lingüístico incluir expresamente a los sistemas de IA dentro del concepto de sistema automatizado o algorítmico, diferenciando simplemente entre si operan con IA o no.

(35) F. GONZÁLEZ CABANES y N. DÍAZ DÍAZ (2023: 40).

Para que un modelo que utiliza técnicas de aprendizaje automático pueda operar de forma eficiente necesita grandes cantidades de datos, de ahí que el procesamiento masivo de datos (*big data*) y la minería de datos (36) (*data mining*) sean imprescindibles en el desarrollo de este campo del conocimiento. Recientemente la Asociación Española de Normalización ha ratificado la norma técnica internacional «Marco para sistemas de inteligencia artificial (IA) que utilizan Machine Learning (ML) (ISO/IEC 23053:2022)» (37). Esta norma de naturaleza voluntaria aporta un conjunto de conceptos comunes y proporciona una base para la explicación estandarizada de esta categoría específica de sistemas de IA, aplicables tanto a la fase de desarrollo como a la fase de comercialización, uso o regulación.

2. Las diferentes técnicas de aprendizaje automático

A) Aprendizaje supervisado (38)

Como su propio nombre indica, el funcionamiento de los sistemas que emplean esta técnica de aprendizaje es siempre supervisado y dirigido por un ser humano. En las fases de entrenamiento y de validación al programa se le proporciona un conjunto de datos etiquetados [datos de entrenamiento (39) y de validación (40)] similares a los que va a procesar una vez se introduzca en el mercado o sea puesto en servicio.

A continuación, a los datos de entrada etiquetados se les predetermina una variable de respuesta (un resultado concreto). De este modo, el modelo va

(36) La minería de datos comprende «la aplicación de algoritmos para el descubrimiento de información válida, novedosa y útil a partir de los datos» (especificación 9.3 de la norma UNE-EN ISO/IEC 22989:2023).

(37) El código español (UNE-EN ISO/IEC 23053:2023) ha sido publicado en el BOE a través de la Resolución de 4 de septiembre de 2023, de la Dirección General de Industria y de la Pequeña y Mediana Empresa, por la que se publica la relación de normas europeas que han sido ratificadas durante el mes de agosto de 2023 como normas españolas.

(38) La especificación 5.11.1 de la norma UNE-EN ISO/IEC 22989:2023 define el aprendizaje supervisado como «el aprendizaje automático que utiliza datos etiquetados durante el entrenamiento».

(39) El art. 3, apartado 29 del Proyecto de Reglamento define los datos de entrenamiento como «los datos usados para entrenar un sistema de IA mediante el ajuste de sus parámetros entrenables, entre los que se incluyen los pesos de una red neuronal».

(40) El art. 3, apartado 30 del Proyecto de Reglamento define los datos de validación como «los datos usados para proporcionar una evaluación del sistema de IA entrenado y adaptar sus parámetros no entrenables y su proceso de aprendizaje, entre otras cosas, para evitar el sobreajuste. El conjunto de datos de validación puede ser un conjunto de datos independiente o formar parte del conjunto de datos de entrenamiento, ya sea como una división fija o variable».

ajustando sus parámetros y una vez que ha sido entrenado y validado podrá funcionar de forma autónoma sin necesidad de contar con datos etiquetados. Un supuesto de aplicaciones que emplean aprendizaje automático supervisado lo encontramos en el filtro de detección de *spam* de nuestro correo electrónico.

B) Aprendizaje no supervisado (41)

En los modelos de IA que utilizan algoritmos de aprendizaje no supervisado los datos de entrada no están etiquetados (en otras palabras, el desarrollador no realiza una labor de clasificación previa y los datos son procesados en bruto). Dado que la variable de respuesta no ha sido predeterminada, no podemos conocer el motivo exacto por el que la máquina llega a un resultado concreto.

La falta de transparencia y explicabilidad de estos modelos hace que su uso deba restringirse en las Administraciones públicas (42). Además, es muy probable que el sistema llegue a un resultado de forma arbitraria y discriminatoria, pues la información de salida se genera mediante correlaciones estadísticas e identificación de patrones comunes desconocidos por el desarrollador o el usuario.

Un ejemplo de la aplicación de técnicas de aprendizaje no supervisado lo encontramos en los sistemas de recomendación que utilizan plataformas como Instagram o Spotify, entre otras.

C) Aprendizaje por refuerzo (43)

La diferencia de esta técnica respecto a las dos anteriores radica en que con este método el modelo aprende a través de la fórmula ensayo-error. Así, en función de cuál sea la variable de respuesta (la información de salida u *output*) se recompensa o penaliza al sistema de IA para que ajuste su fun-

(41) En contraposición al aprendizaje supervisado, el aprendizaje no supervisado se define en la especificación 5.11.2 de la norma UNE-EN ISO/IEC 22989:2023 como el «aprendizaje automático que utiliza datos no etiquetados durante el entrenamiento».

(42) Para profundizar en los problemas jurídicos derivados del uso de sistemas de aprendizaje automático no supervisado en la Administración Pública, véase J. A. TAHIRÍ MORENO, (2023): «El principio de presunción de falibilidad de las decisiones algorítmicas desfavorables: Una nueva garantía jurídica frente a las decisiones automatizadas y el uso de sistemas de inteligencia artificial en la administración pública», *Revista Aragonesa de Administración Pública*, 60, pp. 188-214.

(43) El aprendizaje por refuerzo es el proceso de entrenamiento de un agente o agentes que interactúan con su entorno para alcanzar un objetivo predefinido. En el aprendizaje por refuerzo, los agentes aprenden mediante un proceso iterativo de ensayo y error (especificación 5.11.4 de la norma UNE-EN ISO/IEC 22989:2023).

cionamiento al objetivo deseado. En otras palabras, el sistema incrementa su eficacia planificando la estrategia más óptima para obtener recompensas o estímulos positivos. Este tipo de aprendizaje no es precisamente innovador, ya que se fundamenta en las teorías del condicionamiento operante o instrumental de E. L. THORNDIKE y B. F. SKINNER, aplicadas en el campo de la pedagogía y la psicología.

3. La inteligencia artificial generativa (44)

El boom de la IA generativa (subcampo del aprendizaje automático) tuvo lugar con la puesta en servicio de la aplicación ChatGPT desarrollada por la empresa OpenAI. Sin embargo, no es la única, pues también existen programas informáticos como Midjourney o Deep Voice que crean imágenes y audios con esta técnica.

Lo novedoso de estos modelos es que no se limitan a categorizar, clasificar o recomendar, sino que son capaces de crear contenido nuevo y original (45) (desde arte pictórico hasta guiones de películas (46) o trabajos académicos). Para su funcionamiento utilizan técnicas de aprendizaje automático basadas en el aprendizaje profundo que intentan simular el proceso de aprendizaje del cerebro humano a través de redes neuronales artificiales organizadas en múltiples capas ocultas (47).

Tal es su potencial que en marzo de 2023 el *Future Life Institute* publicó una carta abierta en la que se pedía parar durante 6 meses el entrenamiento de sistemas de inteligencia artificial generativa. La carta, denominada «Pause

(44) Al lector que quiera profundizar en el sector de la inteligencia artificial generativa le resultará de interés la lectura del informe de la OCDE LORENZ, P., K. PERSET and J. BERRYHILL (2023): «Initial policy considerations for generative artificial intelligence», *OECD Artificial Intelligence Papers*, 1, OECD Publishing, Paris, <https://doi.org/10.1787/fae2d1e6-en>.

(45) Aunque la naturaleza original del contenido parece cuestionable desde la perspectiva del derecho de propiedad intelectual. De hecho, son varias las compañías tecnológicas que ya han sido demandadas en Estados Unidos por un posible uso no autorizado de materiales protegidos por derechos de autor para entrenar a sus sistemas de IA. Respecto al Reglamento de IA, parece que finalmente se introducirán obligaciones dirigidas a garantizar el respeto de la normativa europea en materia de derechos de autor.

(46) Precisamente el uso de esta tecnología en la industria cinematográfica estadounidense ha sido una de las causas de la huelga declarada por el Sindicato de Guionistas de Estados Unidos (WGA) y el Sindicato de Actores de Cine-Federación Estadounidense de Artistas de Radio y Televisión (SAG-AFTRA) en mayo y julio de 2023, respectivamente.

(47) Un artículo de acceso libre sobre el funcionamiento del aprendizaje profundo lo encontrarán en *Google Learn* pinchando en el siguiente enlace <https://learn.microsoft.com/es-es/azure/machine-learning/concept-deep-learning-vs-machine-learning?view=azureml-api-2>.

Giant AI Experiments: An Open Letter» (48), advierte de los peligros de un desarrollo descontrolado y piden una regulación que garantice la confianza, la robustez y la transparencia de estos sistemas.

Sin duda, la IA generativa abre una ventana de oportunidades y, con toda probabilidad, transformará a medio plazo la forma en que aprendemos, trabajamos y generamos conocimiento. Sin embargo, es necesario que su uso sea regulado y sometido a determinados requisitos. Para garantizar la protección de la ciudadanía, el texto que recoge las enmiendas del Parlamento Europeo al Proyecto de Reglamento de Inteligencia Artificial incluye deberes específicos de publicidad aplicables a los sistemas de IA generativa, con la finalidad de que los usuarios sean conscientes de que el contenido ha sido generado a través de esta categoría de sistemas (49).

III. LOS RIESGOS Y OPORTUNIDADES DERIVADOS DEL USO DE SISTEMAS DE INTELIGENCIA ARTIFICIAL EN LA ADMINISTRACIÓN PÚBLICA

1. Una exposición de los principales riesgos derivados del uso de sistemas de inteligencia artificial en la actuación administrativa

A) La discriminación algorítmica

El sesgo algorítmico es uno de los principales retos a superar si se quiere implementar de forma generalizada la IA en los procedimientos de actuación de la Administración pública. Evitar situaciones de discriminación provocadas por los sistemas de IA es un objetivo europeo que quedó reflejado en la Declaración Europea sobre los Derechos y Principios Digitales para la Década Digital (2023/C 23/01), de 23 de enero de 2023. En concreto, en virtud de la declaración 9, letra c), el Parlamento, el Consejo y la Comisión adquieren el compromiso de «velar por que los sistemas algorítmicos se basen en conjuntos de datos adecuados para evitar la discriminación y permitir la supervisión

(48) El documento, firmado por miles de expertos del sector tecnológico, puede leerse en <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>.

(49) Establece la enmienda 486 que «Los usuarios de un sistema de IA que genere o manipule contenido de texto, sonido o visual que pueda inducir erróneamente a pensar que tal contenido es auténtico o verídico y que presente representaciones de personas que parecen decir o hacer cosas que no han dicho ni hecho, sin su consentimiento (ultrafalsificación), harán público de manera adecuada, oportuna, clara y visible que el contenido ha sido generado de forma artificial o manipulado, así como, cuando sea posible, el nombre de la persona física o jurídica que lo generó o manipuló. Por hacer público se entenderá etiquetar el contenido de un modo que informe que el contenido no es auténtico y que resulte claramente visible para su destinatario. Para etiquetar el contenido, los usuarios tendrán en cuenta el estado de la técnica generalmente reconocido y las normas y especificaciones armonizadas pertinentes».

humana de todos los resultados que afecten a la seguridad y los derechos fundamentales de las personas».

La discriminación puede tener su origen tanto en la deficiente calidad de los datos de entrada como en las técnicas empleadas por el sistema de IA para hallar correlaciones (50). Si los datos utilizados no son representativos, no están anonimizados o contienen sesgos el modelo puede llegar a resultados discriminatorios. En cuanto a las técnicas de aprendizaje, puede ocurrir (y ocurre con frecuencia) que en el aprendizaje no supervisado se utilicen algoritmos de *clustering* o aproximación [los cuales realizan segmentaciones de datos en función de sus características comunes (51)].

Como botón de muestra, un sistema de IA puede llegar a la conclusión de que las personas que viven en un barrio determinado o forman parte de una minoría étnica son más propensas a cometer fraudes. Aunque fuera cierto desde un punto de vista estadístico, tomar decisiones basadas exclusivamente en datos étnicos o de ubicación geográfica es ilegal e inconstitucional, salvo que se cumplan los requisitos previstos por el apartado 2 del art. 2 de la Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación (52), aplicado en conexión con el apartado 2 del art. 4 de la misma ley (53).

Un supuesto paradigmático de discriminación algorítmica y falta de transparencia en la Administración pública en Europa lo encontramos en el sistema «SyRI», un programa informático utilizado por la Inspección de Trabajo y Seguridad Social holandesa para prevenir y detectar el fraude en la obtención de ayudas sociales. El asunto tuvo repercusión a nivel internacional por la gravedad de los hechos, pues se demostró que la mayoría de los investigados y sancionados eran personas pertenecientes a minorías inmigrantes en riesgo de exclusión social.

La imposibilidad de conocer cómo llega a un resultado el algoritmo unido a la desproporcionalidad de los datos personales utilizados para elaborar los informes que servían para decidir si se iniciaba una investigación llevaron

(50) A. SORIANO ARNANZ (2021: 91).

(51) F. GONZÁLEZ CABANES y N. DÍAZ DÍAZ (2023: 54).

(52) Prevé el precepto citado que «... de acuerdo con lo establecido en el apartado 2 del artículo 4 de esta ley, podrán establecerse diferencias de trato cuando los criterios para tal diferenciación sean razonables y objetivos y lo que se persiga es lograr un propósito legítimo o así venga autorizado por norma con rango de ley, o cuando resulten de disposiciones normativas o decisiones generales de las administraciones públicas destinadas a proteger a las personas, o a grupos de población necesitados de acciones específicas para mejorar sus condiciones de vida o favorecer su incorporación al trabajo o a distintos bienes y servicios esenciales y garantizar el ejercicio de sus derechos y libertades en condiciones de igualdad».

(53) Este apartado establece que «No se considera discriminación la diferencia de trato basada en alguna de las causas previstas en el apartado 1 del artículo 2 de esta ley derivada de una disposición, conducta, acto, criterio o práctica que pueda justificarse objetivamente por una finalidad legítima y como medio adecuado, necesario y proporcionado para alcanzarla».

al órgano judicial a declarar contraria al art. 8 del Convenio Europeo de Derechos Humanos la normativa que regulaba el uso de este programa (54).

B) El sesgo de automatización

Otro riesgo en el que no solemos reparar es el de la confianza excesiva que los ciudadanos depositamos en los resultados generados por las aplicaciones de IA. Un ejemplo claro de este sesgo lo encontramos en los sistemas de vigilancia predictiva empleados por determinados cuerpos de policía en los Estados Unidos. La aplicación «PredPol» analiza en tiempo real el tipo de delito, su ubicación y la fecha/hora (55). De este modo, los agentes cuentan con un mapa de calor que utilizan para distribuir las patrullas de forma más eficiente. A pesar de su eficacia, la utilización acrítica del programa puede tener como consecuencia la estigmatización de núcleos urbanos y la discriminación de determinados colectivos, además de generar lagunas de impunidad en otras ubicaciones donde también se cometen delitos (56).

Junto a las aplicaciones de vigilancia predictiva, un supuesto de sesgo de automatización que quizás le resulte más familiar al lector (por haberlo sufrido en su propia piel alguna vez) es el de los errores humanos provocados por los sistemas de navegación (GPS). En ocasiones es probable que la no actualización de los datos nos haga recorrer más kilómetros o realizar rutas obsoletas. Las muertes por GPS son numerosas y demuestran la fe ciega con la que los seres humanos usamos a veces la tecnología (57). Para evitar situaciones como las descritas es necesario que seamos conscientes de este sesgo y analicemos la información dada por el sistema de IA de forma crítica.

C) La falta de transparencia de los algoritmos

La falta de transparencia de los algoritmos (también conocida como efecto caja negra o *blackbox effect*) es un defecto en el que incurren los modelos de

(54) Un análisis de la sentencia lo encontraréis en L. COTINO HUESO (2020): «SyRI, ¿a quién sanciona? 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*, 4.

(55) Puede accederse a un resumen del funcionamiento de la aplicación a través del siguiente enlace <https://leb.fbi.gov/articles/featured-articles/predictive-policing-using-technology-to-reduce-crime>.

(56) Para una mejor comprensión de los riesgos relativos a la policía predictiva y el Estado vigilante, véase R. RIVERO ORTEGA (2023): «Algoritmos, inteligencia artificial y policía predictiva del Estado vigilante», *Revista General de Derecho Administrativo*, 62.

(57) Recientemente Google ha sido demandada como responsable de la muerte de un hombre que, guiado por la aplicación *Google Maps*, se dispuso a cruzar un puente que llevaba tiempo derrumbado. La noticia, de septiembre de 2023, puede consultarse en <https://computerhoy.com/google/google-demandada-muerte-hombre-seguia-instrucciones-maps-1308510>.

aprendizaje automático no supervisado o de aprendizaje profundo. En virtud del derecho a una buena administración previsto en el art. 41 de la Carta Europea de Derechos Fundamentales (58) y a la exigencia de motivación de los actos administrativos establecida en el art. 35 de la Ley 39/2015 de Procedimiento Administrativo Común (59), las decisiones administrativas deben estar motivadas cuando afecten a los derechos e intereses legítimos del administrado.

Asimismo, que el ciudadano desconozca las razones fácticas y jurídicas de un acto administrativo vulnera el principio de interdicción de la arbitrariedad de los poderes públicos reconocido por el art. 9.3 de nuestra Norma Fundamental. De ahí que, a mi juicio, tenga difícil encaje constitucional y legal la toma de decisiones apoyadas únicamente en el resultado dado por un sistema de IA (60).

En consecuencia, para que la Administración pueda usar un modelo de IA este debe ser transparente y explicable. La transparencia es un principio de actuación de las Administraciones Públicas [art. 3.1.c) LRJSP]. También el art. 11.1.i) del Real Decreto 203/2021, de 30 de marzo, por el que se aprueba el Reglamento de actuación y funcionamiento del sector público por medios electrónicos prevé obligaciones de transparencia en relación con las actuaciones administrativas automatizadas (61).

El estado actual de la técnica nos impide conocer el proceso interno que realiza un sistema de IA que emplea técnicas de aprendizaje automático no supervisado para llegar a un resultado. Esta naturaleza opaca se traduce en una merma de las garantías del ciudadano frente a las actuaciones de la Administración (62). Por otra parte, es probable que aunque conociéramos los motivos de la decisión la actuación administrativa incurriría en vicios de nulidad, pues las técnicas empleadas por esta clase de modelos son por regla general discriminatorias (63).

(58) J. PONCE SOLÉ (2022: 4).

(59) E. GAMERO CASADO (2023b: 17 y 18).

(60) Aquí entraría en juego el principio de «reserva de humanidad» propuesto por PONCE SOLÉ, en virtud del cual el uso de sistemas de IA debería excluirse en determinados ámbitos de actuación de la Administración. Una explicación más detallada del concepto se encuentra en J. PONCE SOLÉ (2022): «Reserva de humanidad y supervisión humana de la inteligencia artificial», *El Cronista del Estado Social y Democrático de Derecho*, 100, pp. 58-67;

(61) Señala el precepto que toda sede electrónica pondrá a disposición de los ciudadanos una «Relación actualizada de las actuaciones administrativas automatizadas vinculadas a los servicios, procedimientos y trámites descritos en la letra anterior (la letra h) se refiere a los servicios, procedimientos y trámites disponibles). Cada una se acompañará de la descripción de su diseño y funcionamiento, los mecanismos de rendición de cuentas y transparencia, así como los datos utilizados en su configuración y aprendizaje».

(62) A. BOIX PALOP y A. SORIANO ARNANZ (2023: 262).

(63) Por ejemplo, el comportamiento de los algoritmos de *clustering* o agrupamiento se ajusta como un guante a la definición de discriminación por asociación prevista en la Ley

2. Las oportunidades y ventajas de su implementación en la Administración Pública

A) Incremento de la eficacia y de la eficiencia en la prestación de los servicios públicos

La lentitud en la tramitación de los procedimientos administrativos y en la prestación de servicios públicos en muchas ocasiones viene motivada por la escasez de recursos humanos. Gracias a los sistemas de IA será posible descongestionar la carga de trabajo de los empleados públicos y dirigir los medios personales de la Administración hacia la realización de tareas que aporten un mayor valor añadido. En este sentido, las Administraciones Públicas deben avanzar hacia un modelo en virtud del cual el grado de automatización de la actividad administrativa sea máximo en las tareas rutinarias de gestión y meramente auxiliar cuando se ejerzan, entre otras, tareas de inspección o control (64).

Pero, no solo se mejora la gestión pública en términos de eficiencia, sino también de eficacia. Un ejemplo cristalino lo encontramos en los asistentes virtuales del Instituto Nacional de la Seguridad Social y de la Agencia Tributaria. Ambos robots conversacionales resuelven consultas a diario, ayudando a los ciudadanos a realizar trámites en la sede electrónica. Todo ello sin tiempos de espera telefónica y sin citas previas.

B) La personalización de los servicios públicos (65)

El análisis masivo de datos y la IA permitirá a la Administración transitar hacia un modelo de gestión administrativa proactivo y personalizado. Así, las

15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación. Así, el art. 6, apartado 2, letra a) dispone que se produce discriminación por asociación «cuando una persona o grupo en que se integra, debido a su relación con otra sobre la que concurre alguna de las causas previstas en el apartado primero del artículo 2 de esta ley, es objeto de un trato discriminatorio». El apartado primero del art. 2, que regula el ámbito subjetivo de aplicación de la norma, establece que «Nadie podrá ser discriminado por razón de nacimiento, origen racial o étnico, sexo, religión, convicción u opinión, edad, discapacidad, orientación o identidad sexual, expresión de género, enfermedad o condición de salud, estado serológico y/o predisposición genética a sufrir patologías y trastornos, lengua, situación socioeconómica, o cualquier otra condición o circunstancia personal o social». Ergo, un algoritmo cuya finalidad es llegar a una conclusión mediante la agrupación de individuos en grupos que compartan características comunes estaría aportando un resultado discriminatorio por asociación.

(64) Una exposición resumida sobre la teoría de los grados de automatización de la actuación administrativa propuesta por Roehl puede encontrarse en L. COTINO HUESO (2023): «Los usos de la inteligencia artificial en el sector público, su variable impacto y categorización jurídica», *Revista Canaria de Administración Pública*, 1, pp. 225.

(65) Un desarrollo del principio de personalización y proactividad en la Administración Pública puede encontrarse en C. VELASCO RICO (2020): «Personalización, proactividad e inteligencia artificial. ¿Un nuevo paradigma para la prestación electrónica de servicios públicos?». *Revista de Internet, Derecho y Política*, 30.

tecnologías emergentes podrían ser utilizadas en un futuro por las Administraciones públicas para anticiparse a las necesidades del ciudadano.

En virtud del principio de proactividad, ya no es el administrado quien formula una solicitud, sino que es la propia Administración quien reconoce de oficio el derecho o la situación jurídica individualizada del ciudadano. No obstante, debemos ser cautos, pues este modelo de gestión no está exento de riesgos y las probabilidades de discriminación son muy altas.

IV. EJEMPLOS DE PROGRAMAS INFORMÁTICOS QUE EMPLEAN INTELIGENCIA ARTIFICIAL EN LA ACTIVIDAD ADMINISTRATIVA

A) La aplicación informática del Cuerpo Nacional de Policía «Veripol». (66)

Esta aplicación informática utilizada por el Cuerpo Nacional de Policía fue desarrollada para detectar denuncias falsas relacionadas con hechos que podrían constituir delitos de robo con fuerza e intimidación. El modelo de IA utiliza técnicas basadas en el aprendizaje automático y su eficacia es del 91%, 15 puntos superior a la de un agente experto (67).

B) El sistema «KENDOJ» del Centro de Documentación Judicial (68)

Creada por el Centro de Documentación Judicial (CENDOJ), esta herramienta permite la pseudonimización automática de sentencias y, además, facilita la lectura de las mismas al enriquecer el texto con enlaces automatizados a la jurisprudencia y legislación citadas en las resoluciones judiciales. No hace muchos años, la mera recopilación de sentencias o la búsqueda de jurisprudencia eran tareas arduas y pesadas. En la actualidad, los motores de búsqueda que operan con IA nos permiten tener en segundos todo el historial de un asunto concreto.

Además, resulta interesante resaltar lo que ya es un hecho notorio: que las editoriales jurídicas llevan tiempo invirtiendo ingentes recursos en aplicaciones de analítica de datos legales (*legal analytics*). Estos programas aportan información estadística sobre cualquier clase de pleito segmentada por jurisdicción, órgano judicial e incluso por jueces y magistrados (69).

(66) <https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/interior/Paginas/2018/271018veripol.aspx>.

(67) L. QUIJANO SÁNCHEZ (2018: 1).

(68) <https://confilegal.com/2021/12/18-la-carrera-judicial-contara-desde-el-lunes-con-una-aplicacion-basada-en-la-inteligencia-artificial-y-machine-learning/>.

(69) Entre otras, Tirant lo Blanch cuenta con la herramienta de analítica jurídica «*Tirant Analytics*» y LA LEY con la aplicación «*Jurimetría*».

C) El asistente virtual «ISSA» del Instituto Nacional de la Seguridad Social

Este robot conversacional (*chatbot*) utilizado por el Instituto Nacional de la Seguridad Social emplea técnicas de IA basadas en el procesamiento de lenguaje natural (70) y sirve de soporte al ciudadano, que puede formular consultas sin necesidad de solicitar una cita previa. El asistente virtual incorpora los 36 servicios más demandados por el ciudadano y las respuestas a las preguntas más frecuentes.

D) La herramienta de valoración del riesgo de la administración penitenciaria catalana «RisCanvi» (71)

Utilizada en el ámbito de la Administración penitenciaria, RisCanvi es una herramienta de valoración del riesgo que lleva operando en Cataluña desde el año 2010. Una de sus finalidades es predecir el grado de probabilidad de reincidencia de los internos (aunque también puede medir la probabilidad de violencia intrainstitucional y autodirigida). La aplicación cuenta con 43 variables o factores de riesgo y recoge el comportamiento individual de cada sujeto.

V. LA REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL A NIVEL EUROPEO Y AUTONÓMICO

1. Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión (72)

La propuesta fue presentada el 21 de abril de 2021 y la tramitación del futuro acto legislativo se encuentra en su fase final. La aprobación definitiva del Reglamento está prevista para el primer trimestre del 2024 y su entrada en vigor es probable que no tenga lugar hasta el año 2026, a fin de que

(70) El procesamiento de lenguaje natural puede definirse como «la disciplina que se ocupa del modo en que los sistemas adquieren, procesan e interpretan el lenguaje natural» (especificación 3.6.10 de la norma UNE-EN ISO/IEC 22989:2023). Por lenguaje natural se entiende cualquier lenguaje humano que pueda ser representado en textos, discursos, lenguaje de signos, etc. Por ejemplo, el idioma español sería un lenguaje natural, en contraposición de Python, que es un lenguaje de programación.

(71) Un resumen sobre el funcionamiento del protocolo Riscanvi lo encontraréis en https://justicia.gencat.cat/web/.content/home/ambits/reinsercio_i_serveis_peni/resum-ris-canvi.pdf (el texto está redactado en catalán).

(72) Puede accederse a la propuesta a través del siguiente enlace <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A52021PC0206>.

los operadores afectados por la norma puedan adaptarse a ella con plenas garantías (73). El pasado 8 de diciembre, tras una intensa negociación (74), se alcanzó un acuerdo provisional sobre el contenido del acto legislativo, finalizando así el periodo de diálogo interinstitucional favorablemente (75). El texto definitivo [que aún no ha sido redactado (76)] deberá ser aprobado por el Parlamento y el Consejo antes de su publicación definitiva en el Diario Oficial de la Unión Europea (77).

El objetivo principal del proyectado Reglamento es garantizar la protección de la salud, la seguridad, los derechos fundamentales reconocidos en cada Estado miembro y los valores de la Unión Europea. Asimismo, el futuro acto legislativo pretende crear las condiciones necesarias para desarrollar un mercado único que evite una situación de fragmentación normativa que genere obstáculos técnicos y dificulte la libre circulación de los sistemas de IA. Para conseguirlo se establecen una serie de requisitos esenciales que deberán

(73) En relación con el grado de cumplimiento actual del futuro Reglamento por parte de las empresas desarrolladoras, son interesantes los resultados del estudio realizado por el *Center for Research on Foundation Models* de la Universidad de Stanford «*Do Foundation Model Providers Comply with the Draft EU AI Act?*». En él se concluye que la aprobación definitiva del Reglamento de IA contribuirá a materializar cambios positivos en el ecosistema de los modelos fundacionales, principalmente en materia de transparencia.

(74) Alemania, Francia e Italia presentaron una propuesta que apostaba por implementar un modelo basado en la autorregulación a través de códigos de conducta, en línea con las reglas acordadas por el Grupo de los 7 (G7) en la «Declaración de los dirigentes del G-7 sobre el proceso de la IA de Hiroshima» publicada el 30 de octubre.

(75) Aunque en el presente epígrafe se comentan algunas novedades relativas al uso de sistemas de IA, es recomendable que las personas interesadas lean los resúmenes del acuerdo publicados en las páginas oficiales del Parlamento y el Consejo Europeo. La nota de prensa del Parlamento podrán encontrarla en el siguiente enlace <https://www.europarl.europa.eu/news/es/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai>. Respecto a la información facilitada por el Consejo, véase <https://www.consilium.europa.eu/es/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/>. Ambos textos están redactados en inglés.

(76) Una vez alcanzado el acuerdo provisional en la fase de trilogos, corresponde a los equipos técnicos de las instituciones implicadas elaborar el texto definitivo. Posteriormente, el documento será revisado por el cuerpo de juristas lingüistas de la Unión Europea antes de su aprobación por el Parlamento y el Consejo Europeo. Por este motivo es muy probable que el Reglamento no esté listo para su adopción definitiva hasta inicios del año 2024.

(77) Entre las novedades principales respecto a la propuesta de Reglamento inicial destacan la obligación de realizar una evaluación del impacto sobre los derechos fundamentales de los sistemas de IA de alto riesgo de forma previa a su comercialización o puesta en servicio; la inclusión de reglas específicas aplicables a los modelos fundacionales (diferenciándose entre modelos fundacionales normales y de «alto impacto»), así como el reconocimiento de situaciones en las que se utilicen sistemas de IA de propósito general, ya sea de forma aislada o integrada en un sistema de IA de alto riesgo.

cumplir obligatoriamente todos los sistemas de IA calificados como de alto riesgo que quieran ser comercializados o puestos en servicio en el mercado interior europeo (78).

En aras de no frenar la innovación y el progreso tecnológico, el proyecto de Reglamento establece un enfoque regulatorio basado en los potenciales riesgos que puedan generar los sistemas de IA (79). Se diferencian, pues, tres categorías: sistemas de riesgo inaceptable, de alto riesgo y de riesgo limitado. Los sistemas de riesgo inaceptable estarán prohibidos y no podrán operar en la Unión Europea; los sistemas de alto riesgo deberán someterse a una evaluación previa de la conformidad en la que se demuestre que el *software* cumple los requisitos esenciales previstos en el Capítulo II del Título III del acto legislativo nuevo enfoque; y, finalmente, los sistemas de riesgo limitado, que no estarán sometidos a las obligaciones previstas para los sistemas de alto riesgo (aunque sí deberán cumplir requisitos específicos de transparencia).

A continuación se expondrá de forma esquemática la tipología de sistemas de IA previstos en la proyectada normativa europea en función de los riesgos que representan.

A) Sistemas prohibidos (de riesgo inaceptable) (80)

a.1. Sistemas de IA que utilicen técnicas subliminales que puedan alterar el comportamiento de un modo que provoque o sea probable que pueda provocar perjuicios físicos o psicológicos en las personas.

a.2. Sistemas de IA que aprovechen las vulnerabilidades de un grupo específico de personas para alterar de manera sustancial el comportamiento de una persona que pertenezca a dicho grupo y provoque o pueda provocar perjuicios físicos o psicológicos a esa o a otra persona.

(78) Un estudio muy interesante sobre los actos legislativos europeos «nuevo enfoque», así como de las normas técnicas armonizadas que los desarrollan lo encontrarán en el libro de V. ÁLVAREZ GARCÍA (2020): *Las normas técnicas armonizadas (Una peculiar fuente del Derecho europeo)*, Iustel, pp. 55-78.

(79) Para un estudio pormenorizado del proyecto de Reglamento de Inteligencia Artificial desde la perspectiva normalizadora y de los controles técnicos, véase V. ÁLVAREZ GARCÍA y J. TAHIRÍ MORENO (2023): «La regulación de la inteligencia artificial en Europa a través de la técnica armonizadora del nuevo enfoque», *Revista General de Derecho Administrativo*, 63.

(80) Artículo 5 del proyecto de Reglamento de Inteligencia Artificial. No obstante, debe tenerse en cuenta que el texto definitivo en el que se plasme el acuerdo provisional interinstitucional incluirá nuevas prohibiciones. Así, el futuro Reglamento prohibirá, además, el uso de sistemas de IA de categorización biométrica que utilicen datos sensibles; los sistemas de reconocimiento de emociones en el lugar de trabajo y en las instituciones educativas; los que impliquen la extracción no dirigida de imágenes faciales de internet o circuitos de televisión cerrados (CCTV) para la creación de bases de datos de reconocimiento facial y los sistemas de IA cuya finalidad sea la manipulación conductual cognitiva.

a.2. Sistemas de crédito o puntuación social (*social scoring systems*) (81).

a.3. Sistemas de identificación biométrica remota en espacios de acceso público con la finalidad de aplicar la ley (salvo en supuestos expresamente tasados y condicionado a determinados requisitos para fines policiales) (82).

B) Sistemas de alto riesgo (83)

El proyecto de Reglamento de IA dedica parte de su articulado a regular los requisitos esenciales que deben de cumplir los sistemas de IA calificados como de alto riesgo (84). En este sentido, el texto diferencia entre dos tipos de sistemas: los *softwares* que se integran en productos que ya están sujetos a regulación armonizada o que formen parte de estos como componentes de seguridad y, por otro lado, los sistemas independientes.

b.1. Sistemas que se integren como componentes de seguridad en productos sujetos a legislación armonizada y sistemas independientes que ya estuvieran sometidos a controles *ex ante* por tercera parte en virtud de dicha legislación (85).

b.2. Sistemas independientes que pertenezcan a uno de los sectores previstos en el anexo III del proyecto de Reglamento. En el ámbito de la Administración pública se incluyen expresamente los sistemas que se utilicen para el acceso y el disfrute de servicios públicos y privados esenciales y sus beneficios; los asuntos relacionados con la aplicación de la ley y la gestión de la migración, el asilo y el control fronterizo, entre otros.

Como novedad respecto del proyecto inicial de la Comisión, el texto definitivo impondrá a los usuarios que sean Administraciones públicas la obli-

(81) Esta categoría de sistema basado en la hipervigilancia (muy similar al Gran Hermano orwelliano de la novela *1984*) lleva unos años utilizándose en la República Popular de China. En función de la puntuación obtenida, los ciudadanos pueden acceder a determinados establecimientos y servicios de transporte público u obtener visados para viajar al extranjero.

(82) Fruto de las negociaciones interinstitucionales los sistemas de identificación biométrica quedarán sujetos a previa autorización judicial y solo podrán ser utilizados en supuestos estrictamente tasados. Como botón de muestra, el acuerdo provisional prevé que el uso de sistemas de identificación biométrica en tiempo real quede restringido a supuestos que requieran la búsqueda de víctimas de determinados tipos de delitos; la prevención de amenazas terroristas y la localización o identificación de personas sospechosas de haber cometido uno de los delitos que se especifiquen en el futuro Reglamento.

(83) Artículo 6 del proyecto de Reglamento de Inteligencia Artificial.

(84) El Capítulo II del Título III del proyecto de Reglamento prevé obligaciones en materia de gestión de riesgos; gobernanza de datos; registros; transparencia y comunicación de información a los usuarios; precisión, solidez y ciberseguridad de los sistemas y vigilancia humana, entre otras.

(85) Aquí se incluyen los productos regulados por la legislación armonizada enumerada en el Anexo II del proyecto de Reglamento.

gación de registrar los sistemas de IA de alto riesgo en la base de datos de la Unión Europea que se creará una vez entre en vigor el Reglamento. Esta base de datos tiene por finalidad incrementar la transparencia facilitando al ciudadano un acceso sencillo y gratuito a las características principales de los sistemas calificados como de alto riesgo.

C) Sistemas de riesgo limitado

Constituyen una categoría residual en la que se integran todos aquellos *softwares* que no están incluidos en las dos categorías anteriores. En el ámbito de los controles técnicos de esta clase de sistemas tendrán un papel determinante los códigos de conducta previstos en el art. 69 del proyecto de Reglamento de Inteligencia Artificial.

Estos documentos tienen naturaleza voluntaria. Sin embargo, su cumplimiento tendrá importantes efectos en el mercado europeo de la IA. Así, las empresas que se sometan a controles voluntarios y obtengan una certificación de conformidad estarán mejor posicionadas que aquellas que no hayan recibido tal certificación.

El art. 69 del proyecto de Reglamento contiene dos tipos de códigos de conducta. Sin ánimo de ser exhaustivo, podemos diferenciar entre los códigos de conducta destinados a promover la aplicación voluntaria de los requisitos esenciales previstos para los sistemas de alto riesgo (art. 69.1); y, por otra parte, los códigos de conducta destinados a promover, entre otros criterios, la aplicación voluntaria de los requisitos relativos a la sostenibilidad ambiental o a la diversidad de los equipos de desarrollo (art. 69.2).

2. Decreto-ley 2/2023, de 8 de marzo, de medidas urgentes de impulso a la inteligencia artificial en Extremadura (86)

La Comunidad Autónoma de Extremadura adoptó el 8 de marzo de 2023 una norma jurídica pionera en España que, a tenor de su preámbulo, tiene por objeto «el establecimiento del marco esencial de las medidas destinadas al apoyo, promoción, impulso y desarrollo de los sistemas de inteligencia artificial en la Comunidad Autónoma de Extremadura, regulando los ámbitos principales de aplicación de ésta».

Más allá de la obligada crítica a la inexistencia del presupuesto habilitante de extraordinaria y urgente necesidad, el Decreto-ley introduce una serie de obligaciones y garantías en el uso de sistemas de IA por parte de la Administración pública autonómica que debemos conocer. Así, el Capítulo IV, denominado

(86) <https://www.boe.es/buscar/act.php?id=BOE-A-2023-8795>.

«Utilización de la inteligencia artificial por parte de la Administración pública autonómica», contiene 3 preceptos (artículos 10 a 12).

El artículo 10 (87) incluye objetivos y prioridades cuyo cumplimiento, en gran medida, no depende de la voluntad del empleado público ni del Poder Ejecutivo, sino de la disponibilidad presupuestaria y del estado de la técnica. El apartado 1 establece un mandato genérico de incorporación de sistemas de IA a las entidades del sector público autonómico extremeño. Por su parte, el apartado 2 prevé el establecimiento de mecanismos para que la utilización de la IA redunde en beneficio de la función de los empleados públicos.

Este apartado llama la atención y ha de ponerse en conexión con el artículo 1, el cual contiene en su apartado 2, letra c) como objetivo del Decreto-ley «Establecer mecanismos para que la utilización por parte de la Administración pública autonómica de Extremadura de sistemas de Inteligencia artificial redunde en beneficio de los empleados públicos». Los poderes públicos deben ser conscientes de que el foco no ha de estar en el empleado público, sino en el ciudadano (88).

La implementación de tecnologías emergentes en las actuaciones administrativas no tiene (o no debería tener) por finalidad beneficiar al empleado público, sino mejorar la prestación del servicio y la eficiencia en la gestión de los recursos públicos. Es decir, con la IA tratamos de satisfacer de la manera más eficaz los intereses generales, en cumplimiento del mandato recogido en el artículo 103.1 de nuestra Norma Fundamental (89).

(87) Dispone el art. 10 lo siguiente: «1. Las entidades a las que se refiere el artículo 2 incorporarán sistemas de inteligencia artificial que favorezcan la eficacia y eficiencia en la gestión de los servicios públicos.

2. Asimismo, establecerán los mecanismos para que la utilización de sistemas de inteligencia artificial redunde en beneficio de la función de los empleados públicos.

3. En todo caso, conforme al apartado 1 del artículo 23 de la Ley 15/2022, de 12 de julio, favorecerán la puesta en marcha de mecanismos para que los algoritmos involucrados en la toma de decisiones que utilice tengan en cuenta criterios de minimización de sesgos, transparencia y rendición de cuentas siempre que sea factible técnicamente. En estos mecanismos se incluirán su diseño y datos de entrenamiento, y abordarán su potencial impacto discriminatorio. Para lograr este fin, se promoverá la realización de evaluaciones de impacto que determinen el posible sesgo discriminatorio».

(88) Así se manifiesta en el punto 8 de la Declaración Europea sobre los Derechos y Principios Digitales para la Década Digital (2023/C 23/01) del Parlamento Europeo, el Consejo y la Comisión adoptada el 23 de enero de 2023, el cual establece que «La inteligencia artificial debe ser un instrumento al servicio de las personas y su fin último debe ser aumentar el bienestar humano».

(89) «La Administración Pública sirve con objetividad los intereses generales y actúa de acuerdo con los principios de eficacia, jerarquía, descentralización, desconcentración y coordinación, con sometimiento pleno a la ley y al Derecho».

Finalmente, el apartado 3, con fundamento en el apartado 1 del artículo 23 de la Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación (90), dispone el favorecimiento de «la puesta en marcha de mecanismos para que los algoritmos involucrados en la toma de decisiones que utilice tengan en cuenta criterios de minimización de sesgos, transparencia y rendición de cuentas siempre que sea factible técnicamente». A continuación, el mismo apartado señala que «se promoverá la realización de evaluaciones de impacto que determinen el posible sesgo discriminatorio».

Como puede observarse a simple vista, el artículo 10 carece de naturaleza imperativa. Se trata de un precepto redactado en clave propositiva o aspiracional. Ello se deduce tanto de la rúbrica del precepto (objetivos y prioridades) como de los verbos empleados en la construcción sintáctica de los enunciados (favorecer, promover...). Nos encontramos, pues, ante un artículo lleno de buenas intenciones cuya exigibilidad es inexistente a menos que se desarrolle reglamentariamente y se concreten las obligaciones de forma expresa.

Mayor relevancia tiene, a mi juicio, el artículo 11, rubricado «*Sistemas de inteligencia artificial en la toma de decisiones*» (91). El apartado 1 del precepto citado reconoce la posibilidad de que el sector público autonómico extremeño adopte actos administrativos a través de sistemas de IA conforme a la Carta de Derechos Digitales del Gobierno de España y la Declaración Europea sobre los Derechos y Principios Digitales para la Década Digital (2023/C 23/01). El apartado 2, por su parte, incluye la obligación de publicar el «mecanismo de decisión, de las prioridades asignadas en el procedimiento de evaluación y de la toma de decisiones, así como de todos los datos que puedan impactar en su contenido».

(90) Señala el art. 23.1 que « 1. En el marco de la Estrategia Nacional de Inteligencia Artificial, de la Carta de Derechos Digitales y de las iniciativas europeas en torno a la Inteligencia Artificial, las administraciones públicas favorecerán la puesta en marcha de mecanismos para que los algoritmos involucrados en la toma de decisiones que se utilicen en las administraciones públicas tengan en cuenta criterios de minimización de sesgos, transparencia y rendición de cuentas, siempre que sea factible técnicamente. En estos mecanismos se incluirán su diseño y datos de entrenamiento, y abordarán su potencial impacto discriminatorio. Para lograr este fin, se promoverá la realización de evaluaciones de impacto que determinen el posible sesgo discriminatorio».

(91) «1. La Administración pública autonómica podrá adoptar actos administrativos mediante sistemas de inteligencia artificial en el marco de un procedimiento administrativo, de acuerdo con los Derechos digitales de la ciudadanía en sus relaciones con las Administraciones Públicas, descritos en la Carta de Derechos Digitales del Gobierno de España y la Declaración Europea sobre los Derechos y Principios Digitales para la Década Digital (2023/C 23/01).

2. Para ello, además de los requisitos previstos en el artículo 41 de la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público, se dará la debida publicidad del mecanismo de decisión, de las prioridades asignadas en el procedimiento de evaluación y de la toma de decisiones, así como de todos los datos que puedan impactar en su contenido».

A diferencia del art. 10, este precepto tiene un gran valor jurídico. Por una parte, el apartado 1 autoriza a las entidades que forman parte del sector público autonómico extremeño a utilizar sistemas de IA para adoptar actos administrativos. Es decir, se otorga a la Administración una habilitación legal previa para operar con modelos de IA (92). Además, el apartado 2 se erige como garantía en favor del administrado, pues establece una obligación de publicidad reforzada de los mecanismos de decisión y los elementos prioritarios tenidos en cuenta en la toma de decisiones que puedan haber influido en el contenido del acto administrativo.

Por último, el artículo 12, denominado «*Garantías para la utilización de la inteligencia artificial en los procedimientos administrativos*», establece dos obligaciones en materia de transparencia (93). El apartado 1 dispone que las normas que regulen los procedimientos administrativos deberán hacer referencia expresa «al impacto del uso de sistemas de inteligencia artificial en la prestación de los servicios públicos que, en su caso, soporten la asistencia en la presentación de solicitudes, declaraciones responsables o comunicaciones, la comprobación o verificación de los requisitos de los interesados, así como la toma de decisiones».

A continuación, el apartado 2 incluye la obligación de dejar constancia en el Inventario de Información Administrativa de las validaciones realizadas por el órgano responsable del procedimiento respecto del proceso diseñado para la adopción de los actos, los riesgos existentes, así como cualquier circunstancia

(92) Esta cuestión tiene plena conexión con el principio de vinculación positiva de las Administraciones públicas a la ley. Así, debemos preguntarnos, ¿puede la Administración emplear sistemas de IA sin habilitación legal previa? Si la respuesta es negativa, ¿es suficiente cobertura legal la dispuesta por los arts. 22 del Reglamento General de Protección de Datos (RGPD); 41 de la LRJSP y 23 de la Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación? Algunos juristas quizás dan por hecho que la Administración puede hacer todo aquello que no está prohibido y consideran que un precepto como este carece de relevancia jurídica. En mi opinión, para que la Administración pueda utilizar la IA en sus procedimientos de actuación, tanto *ad intra*, como *ad extra*, es necesario que exista una cobertura normativa previa. Como ya comenté al principio de este estudio, los modelos de IA son una especie de sistema automatizado y requieren una regulación específica que prevea tales diferencias y autorice a la Administración su uso. Solo así se podrán minimizar los supuestos de vulneración de derechos y evitar abusos en el ejercicio de actuaciones administrativas.

(93) «1. Las normas que regulen los procedimientos administrativos harán referencia expresa al impacto del uso de sistemas de inteligencia artificial en la prestación de los servicios públicos que, en su caso, soporten la asistencia en la presentación de solicitudes, declaraciones responsables o comunicaciones, la comprobación o verificación de los requisitos de los interesados, así como la toma de decisiones.

2. En el Inventario de Información Administrativa, se dejará constancia sobre las validaciones realizadas por el órgano responsable de los procedimientos administrativos respecto del proceso lógico diseñado para la realización de los actos, los riesgos que implica y cualesquiera otros aspectos que garanticen los derechos de los interesados».

que garantice los derechos de los interesados. Ambos apartados constituyen un paso importante (aunque insuficiente) en materia de transparencia.

Finalmente, la materialización de los objetivos plasmados en el Decreto-ley se realizará a través de la Estrategia Extremeña de Inteligencia Artificial [ex Disposición adicional única (94)], documento que tendrá naturaleza de instrumento de planificación y valor reglamentario. Su aprobación corresponde al Consejo de Gobierno de la Junta de Extremadura y la elaboración se delegará a un Grupo de trabajo en el que estarán representados tanto los órganos de las Administración pública autonómica como el sector productivo y empresarial vinculado a la inteligencia artificial.

(94) «1. La Junta de Extremadura elaborará y aprobará una Estrategia Extremeña de Inteligencia Artificial (EIIA) para atender los objetivos, fines y prioridades establecidos en el presente Decreto-ley. Tendrá naturaleza de instrumento de planificación, con valor de norma reglamentaria, y será aprobada por el Consejo de Gobierno.

La Estrategia Extremeña de Inteligencia Artificial tendrá, al menos, los siguientes contenidos:

— Análisis de la situación de partida de la inteligencia artificial en Extremadura. Determinación de las actuaciones a desarrollar con el objetivo de alcanzar las finalidades establecidas en esta norma.

— Identificación de sectores de servicio público y de producción especialmente estratégicos en materia de inteligencia artificial. La Estrategia Extremeña de Inteligencia Artificial, además de los fines, objetivos y prioridades establecidos en el presente Decreto-ley, prestará especial atención al apoyo, desarrollo, impulso y promoción de la inteligencia artificial en los ámbitos educativo y universitario, así como en el ámbito sanitario y servicios sociales.

— Determinación y diseño de experiencias piloto para el desarrollo de sistemas de inteligencia artificial para lo cual se promoverá la colaboración público-privada mediante mecanismos de colaboración entre los centros públicos de investigación y la Universidad con las empresas de sectores especialmente estratégicos en materia de inteligencia artificial.

— Nombramiento de una comisión encargada de su seguimiento y valoración. Su composición será público-privada, estando representados en la misma los diferentes sectores, públicos y privados, que hubieren formado parte del grupo de trabajo encargado de la formulación de la mencionada Estrategia.

2. Para la elaboración de la misma se conformará un Grupo de trabajo del que formarán parte representantes de los órganos de la Administración pública autonómica con competencias en materia de ciencia, tecnología, agenda digital y administración digital, TIC e I+D+I, sin perjuicio de la participación de los órganos con competencias en materia de Sanidad, Educación, Empresa, Empleo, Turismo y Agricultura.

También formarán parte del citado grupo de trabajo representantes del sector productivo y empresarial de la Comunidad Autónoma de Extremadura pertenecientes a sectores o áreas de actividad vinculadas con la inteligencia artificial, como son las empresas del ámbito específicamente tecnológico, del sector agroalimentario, logístico, energético, turístico y cualquier otro ámbito susceptible de implementar sistemas de Inteligencia Artificial.

3. En el plazo de tres meses a la entrada en vigor del presente Decreto-Ley se determinará, mediante Acuerdo de Consejo de Gobierno, la organización y funcionamiento del Grupo de Trabajo previsto en los párrafos anteriores.

4. La Estrategia Extremeña de Inteligencia Artificial será aprobada en el plazo de nueve meses desde la entrada en vigor del presente Decreto-ley mediante Decreto del Consejo de Gobierno, y tendrá la naturaleza de instrumento de planificación».

VI. CONSIDERACIONES FINALES

Como el lector habrá podido comprobar a lo largo de estas páginas, la Cuarta Revolución Industrial está difuminando, lenta pero inexorablemente, las fronteras entre disciplinas científicas que a priori parecían inconciliables. La capacitación y alfabetización en materia de IA es una necesidad, no sólo para los empleados públicos y los operadores jurídicos, sino para la ciudadanía en general. La democratización en el acceso a las tecnologías emergentes implica que los poderes públicos deban intervenir y regular la materia. Pero, no de cualquier modo, pues por primera vez en la Unión Europea la regulación va a desbordar el modelo clásico de intervención, el cual tiene, o mejor dicho, tenía por finalidad principal garantizar la seguridad física de las personas y la unidad del mercado.

Así, los Estados son plenamente conscientes de que la IA no solo compromete la seguridad física del individuo, sino también los valores de la Unión Europea y los derechos fundamentales. En este estudio hemos tenido la oportunidad de conocer qué son las ultrafalsificaciones y cómo las aplicaciones que generan este tipo de contenidos permiten que cualquier individuo sin conocimientos informáticos pueda sustraer atributos de la personalidad como la voz y el rostro. Suplantar nuestra identidad, pues, jamás había sido tan sencillo. También hemos podido comprobar cuán fácil es que un modelo de IA actúe de forma discriminatoria, ya sea por la calidad o la naturaleza de los datos; por la forma de operar de los algoritmos o debido a la existencia de posibles defectos en el ajuste de los parámetros del sistema.

Esta preocupación por la protección de los derechos fundamentales ha quedado reflejada en el acuerdo provisional alcanzado entre el Parlamento, la Comisión y el Consejo Europeo el 8 de diciembre de 2023. Tal es la magnitud del peligro al que se enfrenta la sociedad europea que el futuro Reglamento someterá a los sistemas de IA de alto riesgo a una evaluación sobre su impacto en los derechos fundamentales de forma previa a su comercialización o puesta en servicio.

Asimismo, la necesidad de capacitación en materia de IA que inspiran estas páginas deviene en deber cuando de la Administración pública se trata. Los empleados públicos necesitan una formación transversal que comprenda desde el plano interno y organizativo hasta el externo en sus relaciones con la ciudadanía. La presente (y futura) transformación de los procedimientos de actuación administrativa exige, en consecuencia, que el personal a su servicio cuente con una formación avanzada en competencias digitales. Y ello para minimizar las vulneraciones de derechos a los administrados. Ya hemos tenido la ocasión de conocer en este estudio cómo funcionan los sistemas de IA y cuáles pueden ser las consecuencias de su mal uso en el ejercicio de potestades públicas. No son hipótesis futuristas o relatos de ciencia ficción. La IA es presente.

Desde el ámbito académico creo que aquellos que trabajamos el tema tenemos una especial obligación de participar activamente en la divulgación y transferencia de conocimiento a la sociedad. Este artículo surge de una experiencia personal negativa a la hora de profundizar en el sector de la IA desde la dimensión jurídica. Cuando me sumergí por primera vez en la materia [no hace mucho, a principios del año 2021 con un artículo de HUERGO LORA (95)] me resultó complicado encontrar bibliografía que explicara la cuestión de forma didáctica desde el plano legal y tecnológico. En la fecha actual, por suerte, ya existen obras colectivas «híbridas» excelentes elaboradas por juristas e ingenieros de reconocida competencia.

Aun así, me parecía adecuado elaborar un artículo que permita a los lectores no iniciados en la materia contar con una síntesis de los problemas y conceptos fundamentales de la IA aplicada a la Administración en unas pocas páginas.

Espero sinceramente haber cumplido con el objetivo propuesto.

VII. BIBLIOGRAFÍA

- ÁLVAREZ GARCÍA, Vicente (2010): *Industria*, Madrid, lustel.
- (2020): *Las normas técnicas armonizadas (Una peculiar fuente del Derecho europeo)*, Madrid, lustel, 267 pp.
- ÁLVAREZ GARCÍA, Vicente y TAHIRÍ MORENO, Jesús (2023): «La regulación de la inteligencia artificial en Europa a través de la técnica armonizadora del nuevo enfoque», *Revista General de Derecho Administrativo*, 63.
- BOIX PALOP, Andrés y SORIANO ARNANZ, Alba (2023): «Transparencia y control del uso de la inteligencia artificial por las administraciones públicas», en *Derecho Público de la Inteligencia Artificial*, Zaragoza, Fundación Manuel Giménez Abad, pp. 247-283.
- BOMMASANI, Rishi, *et al.* (2021): «On the opportunities and risks of foundation models», University of Stanford, arXiv preprint arXiv:2108.07258.
- (2023): «Do Foundation Model Providers Comply with the Draft EU AI Act», Stanford, *Stanford Center for Research on Foundation Models*.
- CERRILLO MARTÍNEZ, Agustí (2023): «Lección 21. Actuación automatizada, robotizada e inteligente», *Revista de Derecho Público: Teoría y método*, pp. 525-541.

(95) A. Huelgo Lora (2021): «Regular la inteligencia artificial (en Derecho administrativo)», *El blog de la Revista de Derecho Público: Teoría y Método*.

- COTINO HUESO, Lorenzo (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*, 4.
- (2023): «Los usos de la inteligencia artificial en el sector público, su variable impacto y categorización jurídica», *Revista Canaria De Administración Pública*, 1, pp. 21-242.
- GAMERO CASADO, Eduardo (2023): «Sistemas automatizados de toma de decisiones en el Derecho Administrativo Español», *Revista General de Derecho Administrativo*, 63.
- GAMERO CASADO, Eduardo, et al. (2023): *Inteligencia artificial y sector público. Retos, límites y medios*, Valencia, Tirant lo Blanch.
- GONZÁLEZ CABANES, Francisco y DÍAZ DÍAZ, Norberto (2023): «Capítulo I. ¿Qué es la Inteligencia Artificial?», en GAMERO CASADO, Eduardo (dir.) et al., *Inteligencia artificial y sector público. Retos, límites y medios*, Valencia, Tirant lo Blanch, pp. 37-72.
- LORENZ, Philippe, et al. (2023): «Initial policy considerations for generative artificial intelligence», *OECD Artificial Intelligence Papers*, 1, París, *OECD Publishing*.
- MIR PUIGELAT, Oriol (2023): «La automatización y el uso de algoritmos e inteligencia artificial en derecho administrativo comparado», *Revista General de Derecho Administrativo*, 63.
- PONCE SOLÉ, Juli (2022): «Reserva de humanidad y supervisión humana de la inteligencia artificial», *El Cronista del Estado Social y Democrático de Derecho*, 100, pp. 58-67.
- PONCE SOLÉ, Juli (2022): «Las relaciones entre inteligencia artificial, regulación y ética, con especial atención al sector público», *Revista General de Derecho Administrativo*, 61.
- PRESNO LINERA, Miguel Ángel (2022): *Derechos fundamentales e inteligencia artificial*, Madrid, Marcial Pons.
- QUIJANO SÁNCHEZ, Lara, et al. (2018): «Applying automatic text-based detection of deceptive language to police reports: Extracting behavioral patterns from a multi-step classification model to understand how we lie to the police», *Knowledge-Based Systems*, 149, pp. 155-168.
- RIVERO ORTEGA, Ricardo (2023): «Algoritmos, inteligencia artificial y policía predictiva del Estado vigilante», *Revista General de Derecho Administrativo*, 62.
- SORIANO ARNANZ, Alba (2021): «Decisiones automatizadas: problemas y soluciones jurídicas. Más allá de la protección de datos». *Revista de derecho público: teoría y método*, 3, pp. 85-127.

- TAHIRÍ MORENO, Jesús A. (2023): «El principio de presunción de falibilidad de las decisiones algorítmicas desfavorables: Una nueva garantía jurídica frente a las decisiones automatizadas y el uso de sistemas de inteligencia artificial en la Administración Pública», en esta REVISTA, 60, pp. 188-214.
- VELASCO RICO, Clara Isabel (2020): «Personalización, proactividad e inteligencia artificial. ¿Un nuevo paradigma para la prestación electrónica de servicios públicos?», *Revista de Internet, Derecho y Política (IDP)*, 30.