

Documentación Administrativa, número 12, junio de 2024  
Sección: ARTÍCULOS  
Recibido: 19-06-2024  
Modificado: 19-07-2024  
Aceptado: 22-07-2024  
Publicación anticipada: 09-08-2024  
Publicado: 05-09-2024  
ISSN: 1989-8983 – DOI: <https://doi.org/10.24965/da.11401>  
Páginas: 87-102



Referencia: Parres García, A. (2024). «Gervasio Para Todo» y la gobernanza pública algorítmica. *Documentación Administrativa*, 12, 87-102. <https://doi.org/10.24965/da.11401>

## «Gervasio Para Todo» y la gobernanza pública algorítmica

### «Gervasio Para Todo» and the algorithmic public governance

Parres García, Alejandro

Cabildo de Gran Canaria (España – Spain)

ORCID: <https://orcid.org/0009-0003-6048-2187>

[aparresg@grancanaria.com](mailto:aparresg@grancanaria.com)

#### NOTA BIOGRÁFICA

Funcionario de carrera del Cuerpo Superior de Administradores de la Comunidad Autónoma de Canarias. Licenciado en Derecho, máster en Dirección de Recursos Humanos, entre otra formación de postgrado. Ha desempeñado diversos puestos en la administración autonómica canaria, sobre todo en el ámbito de la Intervención General de la Comunidad Autónoma; y de directivo público, tanto en el Gobierno de Canarias (Director General de Relaciones con la Administración de Justicia, Director General de Modernización y Calidad de los Servicios); como en el Cabildo de Gran Canaria, desempeñando actualmente el puesto de Coordinador Insular del Área de Gobierno de Presidencia, Modernización e Innovación Administrativa. Ha desarrollado numerosa actividad docente en el Instituto Canario de Administración Pública en materias relacionados con la función directiva pública profesional (planificación y dirección estratégica, liderazgo, motivación, etc.). Autor de la monografía: «Manual Básico para la Moderna Gestión Pública» y de diversos artículos doctrinales sobre la ciencia de la Administración.

#### RESUMEN

**Objetivos:** analizar la creciente integración de la inteligencia artificial (IA) en la Administración pública y sus implicaciones. **Metodología:** análisis de fuentes doctrinales y normativas nacionales e internacionales, así como comunicados de organizaciones internacionales que han estudiado la materia. **Resultados:** partiendo de la irrupción de la IA avanzada de propósito general, se realiza una visión sintética de la evolución de los paradigmas de reforma en la Administración pública y del impacto de las tecnologías de la información y comunicación, demostrando cómo de meras herramientas de ayuda inicialmente poco relevantes, las TIC han pasado a convertirse en un factor configurador del propio modelo de gobernanza, al que podríamos denominar como «Gobernanza Pública Algorítmica». **Conclusiones:** los sistemas de IA avanzados, y específicamente la IA generativa, están llamados a transformar la Administración pública desde la óptica de la eficacia y eficiencia, pero su implementación plantea potenciales riesgos como el derivado de la opacidad en la forma en que predetermina las decisiones de los poderes públicos, los posibles sesgos discriminatorios, su posible afectación desfavorable sobre los derechos fundamentales de las personas y sobre la sostenibilidad medioambiental, así como una gran incertidumbre sobre cómo debe evolucionar el Derecho público para afrontar los retos que esta nueva «racionalidad algorítmica» plantea.

#### PALABRAS CLAVE

Inteligencia artificial; algoritmos; administración pública; gobernanza pública; innovación pública; valor público; normativa.

#### ABSTRACT

**Objectives:** To analyze the growing integration of artificial intelligence (AI) in public administration and its implications. **Methodology:** Analysis of national and international doctrinal and regulatory sources, as well

as statements from international organizations that have studied the subject. **Results:** Starting from the advent of advanced general-purpose AI, a synthetic view of the evolution of public administration reform paradigms and the impact of information and communication technologies is provided, demonstrating how ICTs, initially mere tools of little relevance, have become a shaping factor of the governance model itself, which we could call «Algorithmic Public Governance». **Conclusions:** Advanced AI systems, and specifically generative AI, are poised to transform public administration from the perspective of efficiency and effectiveness, but their implementation poses potential risks such as the opacity in how it predetermines public decision-making, possible discriminatory biases, its potential adverse impact on fundamental human rights and environmental sustainability, as well as great uncertainty about how public law should evolve to address the challenges posed by this new «algorithmic rationality».

## KEYWORDS

Artificial intelligence; algorithms; public administration; public governance; public innovation; public value; regulation.

## SUMARIO

1. INTRODUCCIÓN: «GERVASIO PARA TODO» Y LOS SISTEMAS DE APRENDIZAJE AUTOMÁTICO PROFUNDO. 2. LA EVOLUCIÓN DE LOS PARADIGMAS DE REFORMA DEL SECTOR PÚBLICO Y EL IMPACTO DE LA TECNOLOGÍA EN EL MISMO. 2.1. LA RACIONALIDAD JURÍDICA DEL MODELO BUROCRÁTICO. 2.2. LA RACIONALIDAD GERENCIAL DE LA NUEVA GESTIÓN PÚBLICA. 2.3. LA RACIONALIDAD RELACIONAL DE LA GOBERNANZA PÚBLICA. 2.4. LA RACIONALIDAD COLABORATIVA PROPIA DEL GOBIERNO ABIERTO Y LA GOBERNANZA PÚBLICA INTELIGENTE. 3. ALGUNAS INCERTIDUMBRES, RIESGOS Y DESAFÍOS, QUE PLANTEA EL USO DE «GERVASIO PT» POR LA ADMINISTRACIÓN PÚBLICA. 3.1. PARA LA LEGITIMIDAD DE LAS RESOLUCIONES ADMINISTRATIVAS PREDETERMINADAS POR LA IA Y SU CONTROL DEMOCRÁTICO. 3.2. LA POSIBLE AFECCIÓN A LOS DERECHOS FUNDAMENTALES DE LA CIUDADANÍA POR ERRORES, SESGOS Y DISCRIMINACIONES ALGORÍTMICAS. 3.3. PARA LA PROTECCIÓN DE LOS DERECHOS DE AUTOR Y DE LOS DATOS PERSONALES. 3.4. LA NECESARIA SUJECCIÓN DE LA IA A PRINCIPIOS ÉTICOS. 4. EL ROL A DESEMPEÑAR POR LA POLÍTICA Y LA ADMINISTRACIÓN PÚBLICA ANTE LOS POSIBLES IMPACTOS DESFAVORABLES DE LA IA SOBRE LA SOCIEDAD. 4.1. SOBRE EL EMPLEO. 4.2. SOBRE LA MOVILIDAD. 4.3. PARA LA SOSTENIBILIDAD MEDIOAMBIENTAL. 4.3.1. Consumo de energía de los Centros de Datos. 4.3.2. Residuos electrónicos. 4.3.3. Impacto del ciclo de vida de los dispositivos. 5. ALGUNAS REFLEXIONES PARA CONCLUIR. REFERENCIAS BIBLIOGRÁFICAS.

## 1. INTRODUCCIÓN: «GERVASIO PARA TODO» Y LOS SISTEMAS DE APRENDIZAJE AUTOMÁTICO PROFUNDO

Ha llegado un asistente nuevo a la oficina, pero no es uno cualquiera. Está causando una auténtica revolución. Yo lo llamo «Gervasio Para Todo», o de forma abreviada GPT (porque lo de «Generative Pre-trained Transformer», que parece ser su nombre verdadero, me parece un esnobismo).

Gervasio parece poseer un conocimiento más que extraordinario de cualquier materia sobre la que le pidas asistencia. Maneja cientos de miles de millones de datos de distintas fuentes (libros, sitios web, artículos), sobre los que identifica patrones para generar sus respuestas a tus demandas en cuestión de segundos.

¡Sí, Gervasio es capaz de hacer magia! Por ejemplo, puedes pedirle que seleccione –de una base de datos abiertos con cientos de miles de documentos sobre contratación administrativa– cinco o diez modelos de pliegos de cláusulas administrativas particulares de contratos cuyo objeto guarde semejanza con el de la necesidad de contratación que debes poner en marcha, y, a continuación, pedirle que los analice y redacte un nuevo borrador de pliego de cláusulas partiendo de información de aquellos, pero personalizándolo para ajustarlo a las características que le señales. Hará en unos minutos lo que hasta ahora podría llevar varias jornadas de trabajo. Si eres hábil para formularle la pregunta o instrucción (prompt o texto inicial de entrada) de lo que quieres obtener (salida) dándole adecuado contexto, será enormemente eficaz y eficiente. ¡Encima es multilingüe!

Hace todo eso y más, pese a que sus condiciones laborales son de auténtica esclavitud: solo hay que pagarle en torno a 20 euros al mes, trabaja 24 horas al día los 7 días de la semana y siempre está dispuesto a atenderte con amabilidad, aunque se lo pidas a las 3 de la madrugada. Gervasio no tiene días de vacaciones, ni permisos por asuntos personales, no se pondrá en huelga, y le resbala las medidas de prevención de la salud o seguridad que pueda haber en el lugar de trabajo.

Algunas de las personas que trabajan en la oficina habían oído hablar de Gervasio, pero no lo conocían. Cuando se les presentó pasaron de la curiosidad al asombro, y del asombro a pedir cada una de ellas un Gervasio como asistente particular.

Pese a la broma de humanizar a Chat GPT, lo que relato viene a representar una pequeña muestra del extraordinario impacto que para la productividad y eficiencia ya supone el uso de herramientas de apoyo basadas en Inteligencia Artificial generativa o de propósito general, que utilizan el procesamiento del lenguaje natural.

La IA ya existía desde mediados de los años 50, pero en los últimos años la evolución de los tres insumos técnicos para producir capacidad de Inteligencia artificial (datos, computación y algoritmos), hace que ya nada vuelva a ser igual y que la IA esté en la cresta de la ola del debate público:

1. El crecimiento de la cantidad de datos digitalizados en internet es enorme: Se estima que en 2025 habrá 175 Zettabytes (Reinsel *et al.*, 2017). Un Zettabyte equivale a mil millones de Terabytes. Un Terabyte almacena aproximadamente 6,5 millones de documentos de texto o alrededor de 25 millones de páginas de Word.
2. El desarrollo extraordinario de la capacidad de cómputo y procesamiento de datos de las TIC ha sido exponencial en las últimas décadas, impulsado por avances tecnológicos en hardware, arquitectura de procesadores y nuevas formas de computación (computación en la nube o la computación cuántica aún en sus primeras etapas).
3. Los algoritmos de aprendizaje automático. El desarrollo vertiginoso de la IA, especialmente de la IA Generativa, un tipo de algoritmo de aprendizaje automático diseñado para comprender el contexto, generar respuestas, prestar asistencia, hacer predicciones, etc. Estos modelos son entrenados en grandes volúmenes de datos y pueden realizar una amplia variedad de tareas relacionadas con el procesamiento del lenguaje natural.

Los algoritmos tradicionales no son otra cosa que una secuencia bien definida y ordenada de pasos o instrucciones para realizar una tarea determinada o resolver un problema específico o alcanzar un objetivo. Es, esencialmente, una receta o un plan de acción. Son la base sobre la cual se construyen las aplicaciones y sistemas, permitiendo a las máquinas realizar tareas complejas de manera rápida y precisa.

Hasta ahora ya disponíamos en la Administración pública de automatización de tareas burocráticas, actos administrativos automatizados de tipo clásico, sin uso de IA, mediante algoritmos estáticos.

Lo que ocurre ahora es que los desarrollos de los sistemas algorítmicos avanzados son dinámicos o predictivos; sistemas diseñados para funcionar con diversos niveles de autonomía y capaces de generar información de salida (diverso contenido, como predicciones, recomendaciones o decisiones) a partir de peticiones usando el lenguaje natural; pudiendo distinguirse dentro de esta nueva tipología a los de aprendizaje automático o *machine learning*, y los de aprendizaje profundo o *deep learning*.

Los de *machine learning* son aquellos que aprenden de forma autónoma a realizar una tarea o hacer predicciones a partir de datos y mejoran su rendimiento con el tiempo. Una vez entrenado, el algoritmo podrá encontrar los patrones en nuevos datos, proporcionando soluciones que difieren de las inicialmente programadas. Los algoritmos de *machine learning* aprovechan los datos estructurados y «etiquetados» para realizar predicciones, lo que conlleva la definición por humanos de características específicas durante la entrada de datos para el modelo y su organización en tablas. Estos algoritmos de machine learning son susceptibles de proporcionar cierta trazabilidad.

De otro lado, los de aprendizaje profundo o *deep learning*, intentan imitar las acciones del cerebro humano creando modelos complejos mediante redes neuronales artificiales interconectadas, con múltiples capas, a partir de datos masivos, y cuya trazabilidad es casi inexistente (Velasco Rico, 2024).

Los sistemas de aprendizaje profundo pueden ingerir y procesar datos no estructurados, como texto e imágenes, y automatizan la extracción de características, eliminando parte de la dependencia de expertos humanos. Cada red neuronal incluye una capa de entrada, varias capas ocultas y una capa de salida. Con cada capa de neuronas oculta el algoritmo procesa los datos recibidos de la capa

anterior y extrae características más complejas de los mismos, ajustándose y adaptándose para ganar precisión. El término «profundo» hace referencia a la gran cantidad de capas de neuronas ocultas en la red neuronal.

«Gervasio Para Todo» emplea este tipo de red neuronal artificial que trata de imitar el funcionamiento del cerebro humano; por lo que no sabemos explicar cómo sabe lo que nos reporta.

## 2. LA EVOLUCIÓN DE LOS PARADIGMAS DE REFORMA DEL SECTOR PÚBLICO Y EL IMPACTO DE LA TECNOLOGÍA EN EL MISMO

A medida que las potencialidades de «Gervasio Para Todo», de la incorporación de la IA generativa (GenAI), en la cadena de creación de valor público se va haciendo más perceptible, comienza a vislumbrarse el surgimiento de un nuevo paradigma de la Gobernanza Pública, al que, por ponerle un nombre, podríamos denominarlo como «Gobernanza Pública Algorítmica», por el omnipresente papel que los algoritmos van a desempeñar en la propia configuración de la misma.

Esta «Gobernanza Pública Algorítmica» comienza a abrir intensos debates públicos en torno a la manera de afrontar las profundas implicaciones que ya se atisban en el horizonte, buscando, por un lado, cómo aprovechar los enormes beneficios que la IA avanzada puede aportar para la generación de valor público, y, por otro, cómo eliminar o mitigar sus posibles efectos desfavorables asociados a los riesgos de que:

- a) Su aplicación por la Administración pública pudiera lesionar garantías o derechos fundamentales de la ciudadanía, dar lugar a decisiones enfrentadas con la ética, produzcan inequidad en la prestación de servicios públicos, o afecten a la protección de la privacidad de los usuarios; además de las grandes incertidumbres que plantea su legitimidad para predeterminar las resoluciones de los decisores públicos por los déficits que plantea en trazabilidad y transparencia.
- b) Su implementación en todos los ámbitos de la vida provoque profundos cambios en la sociedad y economía que den lugar a efectos indeseados (situaciones de discriminación, de desempleo, etc.) a los que los Gobiernos deban dar respuestas, en muchos casos inéditas, pese a no ser causantes de dichas transformaciones.
- c) El efecto negativo que para la sostenibilidad medioambiental global tiene el uso masivo de la IA, a reservas de que se avance en el desarrollo de algoritmos de mayor eficiencia.

Pero, para comprender la dimensión de los desafíos que van a plantear los sistemas algorítmicos avanzados, y del rol que tendrá que desempeñar la Política y la Administración pública para afrontarlos, interesa hacer un repaso muy sintético de la evolución de los paradigmas de reforma que históricamente han intentado aplicarse para modernizar la Administración pública, y ponerlo en contexto con el papel que han tenido las tecnologías de la información y comunicación (TIC, en adelante).

Veremos que el papel de estas últimas ha basculado desde su aparición como mera herramienta para realizar operaciones de cálculo masivos, a pasar a ser un elemento configurador del propio «sistema operativo» de la gobernanza pública.

### 2.1. La racionalidad jurídica del modelo burocrático

Hasta la mitad de los años 70 del siglo pasado, la Administración pública tradicional, basada en el modelo burocrático clásico o Weberiano, alcanza su momento de auge. La racionalidad o lógica que se sostenía como ideal del modelo para funcionar es la jurídica, basada en el principio de legalidad. Las decisiones son racionales, justas, si se toman siguiendo los procedimientos previstos en las leyes y reglamentos, en los que se predetermina el resultado que tendrá la resolución. El Derecho Público trata de evitar la discrecionalidad a toda costa, buscando seguridad jurídica, partiendo de la base de que el legislador ha sido capaz de prever en las leyes la manera de afrontar la problemática, expectativa o desafío al que se dirigen las mismas. En caso de cambios en el contexto o de aparición de desafíos o problemas no previstos en las normas, habrá que modificarlas o elaborar nuevas normas (leyes y reglamentos) para que éstas determinen la ruta que debe seguir el aparato administrativo para tomar decisiones.

En el sector público, guiado por procesos manuales y mecanográficos, las TIC comienzan a introducirse tímidamente, mediante el uso de *mainframes* muy costosos y de uso limitado a tareas específicas, como la elaboración de censos de población o para la gestión de impuestos.

## 2.2. La racionalidad gerencial de la Nueva Gestión Pública

Desde mediados de los 70, el modelo burocrático sobre el que se desarrollan las principales políticas del Estado del bienestar europeo, comienza a ser objeto de profundas críticas, a medida que el contexto de cambios se va acelerando y se reclama que el sector público abandone su lentitud de respuesta (la necesidad de elaboración de leyes y reglamentos para afrontar los sucesivos cambios en el entorno no previstos en las normas vigentes no parece el proceso más ágil para responder con celeridad), su excesiva burocratización, así como la aplicación con fuerza de los principios de eficacia y eficiencia propios del modelo gerencial a la gestión pública, ante la fuerte expansión del gasto público y el impacto de la crisis económica de entonces.

Es la época del paradigma de la «Nueva Gestión Pública»; desarrollándose fundamentalmente durante los años 80 y 90, con distinta intensidad y relativo éxito según los países en los que se impulsa. La racionalidad gerencial es el paradigma de reforma entonces: Para que las decisiones se juzguen racionales no basta con que sean tomadas siguiendo un procedimiento legal o reglamentario, sino que además deben buscar la eficacia y eficiencia.

En esos años comienza a generalizarse la integración de la informática en los servicios públicos, gracias al desarrollo de la microinformática y el despliegue de los ordenadores personales, buscando introducir eficiencia y automatización en los procesos de gestión. Comienzan a digitalizarse los procesos administrativos y a desarrollar sistemas propietarios para funciones básicas, como, por ejemplo, la administración tributaria o la contabilidad pública (mandando a los museos a los grandes libros de contabilidad en los que se hacían los apuntes de forma manual y que median cerca de 1 metro de anchura).

La generalización del uso de Internet desde mediados de los años 90 viene a suponer un momento disruptivo en la historia de la humanidad, que acelerará procesos de cambio social de manera exponencial. Esta red de redes permitirá el desarrollo del fenómeno de la globalización, para referirse al conjunto de procesos interrelacionados que se manifestaban a escala mundial<sup>1</sup>.

## 2.3. La racionalidad relacional de la Gobernanza Pública

Desde el inicio de la primera década del 2000, los efectos no deseados del fenómeno de la Globalización y la creciente desafección ciudadana con la política, llevan a la reflexión de la Comisión Europea en torno a cómo reaccionar ante la creciente percepción ciudadana sobre la pérdida de eficacia y legitimidad de lo público (Comisión Europea, 2001).

Se alcanza cierto consenso en acuñar un nuevo paradigma de reforma al que denomina «Gobernanza Pública», que, sin abandonar la contribución de los paradigmas anteriores (burocracia y gerencia pública) plantea, en síntesis, la necesidad de mejorar la calidad de la interacción entre los distintos niveles de gobierno, las organizaciones empresariales y la sociedad civil; de desarrollar una lógica de gobierno relacional o en redes de interacción público-privada a lo largo del eje local-global. Se parte de la base de que ya no debe ser monopolio de los agentes gubernamentales la definición de las políticas públicas, dado que los Gobiernos no tienen todo el conocimiento sobre los desafíos o expectativas que afectan a la sociedad. Por consiguiente, para «Gobernanza Pública» es preciso incorporar también el conocimiento colectivo de las organizaciones y de la sociedad civil en espacios públicos participativos y deliberativos, en los que se aprenda de la interacción en red para poder tomar decisiones más legitimadas y eficaces (Prats i Catalá, 2005).

En el ámbito TIC es la época del *E-Government* (Gobierno electrónico) impulsando con fuerza la digitalización de procedimientos y servicios públicos; se desarrollan los portales web para que los ciudadanos puedan realizar servicios en línea y comienzan a desarrollarse plataformas que permitan sistemas interoperables, de intercambio de datos entre departamentos y distintos niveles de gobierno<sup>2</sup>.

<sup>1</sup> La Ley 30/1992, de 26 de noviembre, de Régimen Jurídico de las Administraciones Públicas y del Procedimiento Administrativo Común (LRJAP-PAC), en su primera versión, recogió ya en su art. 45 el impulso al empleo y aplicación de las técnicas y medios electrónicos, informáticos y telemáticos, por parte de la Administración al objeto de desarrollar su actividad y el ejercicio de sus competencias y de permitir a los ciudadanos relacionarse con las Administraciones cuando fuese compatible con los medios técnicos de que dispongan.

<sup>2</sup> Bajo el impulso, sobre todo, de la Ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los Servicios Públicos, y su normativa de desarrollo.



## 2.4. La racionalidad colaborativa propia del Gobierno Abierto y la Gobernanza Pública Inteligente

En la segunda década del 2000, este paradigma de reforma, de génesis europea, bascula y se extiende a nivel mundial hacia el paradigma denominado de «Gobierno Abierto», en los años siguientes, popularizado a partir del Memorando emitido por Barack Obama en 2009 a sus Agencias y jefes de Departamentos Ejecutivos, en el que expresaba que la forma de gobernar debía estar presidida por los principios de transparencia, participación y colaboración.

La racionalidad relacional instaurada en el paradigma de la «Gobernanza Pública» adquiere ahora, con «Gobierno Abierto», un barniz tecnológico importante: Las TIC van a posibilitar cada vez con mayor facilidad y con menor coste la captación del conocimiento colectivo del exterior para, combinándolo con el proveniente del interior del sector público, alumbrar decisiones más racionales, fundamentadas en procesos de co-creación, dando lugar a una racionalidad colaborativa.

El paradigma de «Gobierno Abierto» va a apoyarse cada vez más en el vertiginoso desarrollo de la Sociedad de la Información. Aparecen la web 2.0, las plataformas de participación, aplicaciones de gestión de conocimiento colaborativo en red, y se avanza hacia una transformación digital más profunda de los servicios públicos, integrando tecnologías avanzadas como la nube, el big data o el desarrollo de aplicaciones móviles para facilitar la accesibilidad a los servicios públicos. A medida que las TIC contribuyen a mejorar el conocimiento disponible (captado tanto del interior como del exterior) para la toma de decisiones «más inteligentes», y la creación de valor público, «Gobierno abierto» va a ir complementando su denominación con la de «Gobernanza Pública Inteligente».

Y en eso estábamos, hasta que recientemente llegó «Gervasio Para Todo» o el asistente Chat GPT, dando lugar a otro momento disruptivo, posibilitando la generalización del uso e integración de sistemas basados en el aprendizaje profundo a los que nos hemos referido al principio de este artículo; dando lugar a una carrera vertiginosa entre las compañías tecnológicas (fundamentalmente estadounidenses y chinas) para no quedarse atrás en los avances (casi diarios) de esta tecnología. Esta aceleración tecnológica; a su vez, complica la posibilidad de hacer un análisis pausado del alcance y los efectos de su aplicación por la Administración Pública, que se siente impelida a incorporar la innovación que estos avances plantean.

## 3. ALGUNAS INCERTIDUMBRES, RIESGOS Y DESAFÍOS, QUE PLANTEA EL USO DE «GERVASIO PT» POR LA ADMINISTRACIÓN PÚBLICA

### 3.1. Para la legitimidad de las resoluciones administrativas predeterminadas por la IA y su control democrático

Si uno de los principios básicos de la Administración pública, como decíamos en el apartado anterior, es el de la seguridad jurídica, la predeterminación de las decisiones administrativas siguiendo los procedimientos previstos en las normas, estaremos de acuerdo en que el Derecho Público se enfrenta ahora a un enorme desafío, ante la irrupción de sistemas de IA avanzada que van a predeterminar de manera automatizada cual debe ser el contenido de la decisión a adoptar, a través del análisis y procesamiento de macrodatos, basada en cálculos probabilísticos e inferencias algorítmicas, opacas a los decisores públicos, e incluso, llegado un punto, a los propios programadores, generando el efecto que la doctrina científica denominan de «caja negra» de la IA, y sin que dispongamos de un marco jurídico que discipline su uso.

¿Esta nueva racionalidad algorítmica, poco transparente y difícil de explicar, puede ser considerada legítima «a ciegas» por los usuarios de los servicios públicos, por la ciudadanía?, ¿Los empleados públicos que obtengan de la IA generativa contenido que, por ejemplo, se incorpore a informes que avalen las decisiones administrativas, deberán trasladarlo sin más a resoluciones administrativas?

Parece un contrasentido que la humanidad dedique tantos esfuerzos a investigar el funcionamiento del cerebro humano y, sin embargo, dejemos a los sistemas algorítmicos avanzados predeterminar el sentido de una decisión gubernamental sin que comprendamos del todo cómo lo ha hecho.

El otro día le subí a Gervasio un archivo con un borrador de propuesta de modificación de un contrato administrativo, y le pedí que me preparara un informe sobre si se ajustaba a la Ley de Contratos del Sector Público. Se trataba de una modificación no prevista en el Pliego de Cláusulas Administrativas Particulares y teníamos cierto debate técnico interno acerca de si la tramitación de dicha modificación se ajustaba a los requisitos que la LCSP prevé para dar curso a esta clase de modificaciones.

Gervasio, en cuestión de segundos, elaboró un informe muy bien estructurado, que finalizaba de la siguiente manera:

La Propuesta de Modificación de Contrato ha sido revisada y se concluye que cumple con los requisitos establecidos por la Ley de Contratos del Sector Público. Las modificaciones propuestas están justificadas, son transparentes y no alteran las condiciones esenciales del contrato de manera que afecten la competencia. Se han respetado los principios de igualdad de trato y no discriminación, así como las condiciones económicas permitidas por la normativa vigente.

¡Asombroso! Gervasio logró disipar las dudas.

Un rato más tarde, desconfiado –puede que por mi pasado en la Intervención– volví a pedirle a Gervasio que me elaborase de nuevo un informe, pero, esta vez, a diferencia del anterior, que motivase cómo la Propuesta de modificación NO se ajustaba a la LCSP.

En dos segundos Gervasio elaboró otro Informe, igualmente bien estructurado, que finalizaba señalando:

La Propuesta de Modificación de Contrato” ha sido revisada y se concluye que no cumple con los requisitos establecidos por la Ley de Contratos del Sector Público. Las modificaciones propuestas carecen de justificación adecuada, no respetan los principios de transparencia, igualdad de trato y no discriminación, y no se ajustan a las condiciones económicas permitidas por la LCSP.

Desconcertado, le pregunté a Gervasio ¿cuál de los dos informes que había elaborado era el más correcto?, a lo que me contestó lo siguiente:

Para determinar cuál de los dos informes es más correcto en cuanto a la adecuación de la Propuesta de Modificación de Contrato a la Ley de Contratos del Sector Público (LCSP), es esencial revisar el contenido del documento específico y compararlo con los requisitos y principios establecidos en la LCSP. Dado que no tengo la capacidad de interpretar el contenido exacto del archivo de manera directa, a continuación, se presentan los pasos y criterios que debes seguir para evaluar cuál de los informes se ajusta mejor a la realidad del documento:

(...)

Para determinar cuál informe es el más correcto, debes revisar el contenido del documento Propuesta de Modificación de Contrato remitida en detalle, siguiendo los pasos y criterios mencionados anteriormente. Solo una revisión detallada y objetiva del documento permitirá concluir si cumple o no con los requisitos establecidos por la Ley de Contratos del Sector Público. Si necesitas ayuda para analizar un aspecto específico del documento o tienes alguna pregunta adicional sobre cómo interpretar ciertos puntos, estaré encantado de ayudarte.

*¡El conejo me (des)riscó la perra!*, como diría algún pastor de las cumbres de Gran Canaria: Gervasio PT me pide que revise su informe con detalle y objetividad<sup>3</sup>.

El ejemplo anterior, nos puede servir para juzgar el auténtico valor a día de hoy de herramientas de IA generativa como Gervasio; cuyo uso como «abrelatas» de conocimientos es extraordinario, pero siempre debiendo someter sus resultados a refinamiento mediante la aplicación de medidas de comprobación y verificación humana.

Parece obvio que los algoritmos que se usan para la toma de decisiones automatizadas deban ser públicos, auditables y revisables, aunque el ordenamiento jurídico español haya ido disminuyendo las garantías previstas en el derogado art. 45 de la Ley 30/1992, de Régimen Jurídico de las Administraciones Públicas y del Procedimiento Administrativo Común, que preveía, en su aptdo. 4, que «los programas y aplicaciones electrónicos, informáticos y telemáticos que vayan a ser utilizados por las Administraciones Públicas para el ejercicio de sus potestades habrán de ser previamente aprobados por el órgano competente, quien deberá difundir públicamente sus características»; lo que permitía, por ejemplo, exigir dar publicidad al código fuente, si el programa o aplicación dispusiera de algoritmo que predeterminara la decisión administrativa (Boix Palop, 2020).

<sup>3</sup> En el caso de que algún compañero de Intervención del Cabildo de Gran Canaria lea este artículo, que tenga presente que la modificación del contrato fue finalmente tramitada, tras revisarse detalladamente y de forma objetiva su contenido y alcanzarse la conclusión de su absoluta regularidad jurídica.

En la vigente Ley 40/2015, de Régimen Jurídico del Sector Público, buscando favorecer el empleo de medios tecnológicos en el ámbito público, únicamente establece, en su art. 41.2, la posibilidad de auditar el sistema de información y de su código fuente, en caso de que así se decida por el órgano competente a la hora de definir previamente las especificaciones, programación, supervisión y control de calidad, de la actuación administrativa automatizada. Con toda probabilidad, este artículo estaba pensado para los actos administrativos automatizados clásicos, sin IA. La irrupción de los sistemas algorítmicos avanzados y su creciente extensión en la Administración se está efectuando sin que las Leyes de procedimiento administrativo común y de régimen jurídico del sector público, de 2015, se hayan actualizado para dar adecuada respuesta a la problemática que presentan los sistemas algorítmicos avanzados.

La Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación, contiene un precepto (art. 23) destinado a la IA y mecanismos de toma de decisiones automatizadas, en el que se señala:

1. En el marco de la Estrategia Nacional de Inteligencia Artificial, de la Carta de Derechos Digitales y de las iniciativas europeas en torno a la Inteligencia Artificial, *las administraciones públicas favorecerán la puesta en marcha de mecanismos para que los algoritmos involucrados en la toma de decisiones que se utilicen en las administraciones públicas tengan en cuenta criterios de minimización de sesgos, transparencia y rendición de cuentas, siempre que sea factible técnicamente*. En estos mecanismos se incluirán su diseño y datos de entrenamiento, y abordarán su potencial impacto discriminatorio. Para lograr este fin, se promoverá la realización de evaluaciones de impacto que determinen el posible sesgo discriminatorio.
2. Las administraciones públicas, en el marco de sus competencias en el ámbito de los algoritmos involucrados en procesos de toma de decisiones, *priorizarán la transparencia en el diseño y la implementación y la capacidad de interpretación de las decisiones adoptadas* por los mismos.
3. Las administraciones públicas y las empresas *promoverán el uso de una Inteligencia Artificial ética, confiable y respetuosa con los derechos fundamentales*, siguiendo especialmente las recomendaciones de la Unión Europea en este sentido.
4. Se promoverá un sello de calidad de los algoritmos.

Las determinaciones de los citados artículos son enormemente abiertas, flexibles y voluntaristas («favorecerán, priorizarán, promoverán»), subordinando tener en cuenta criterios de minimización de sesgos, transparencia y rendición de cuentas a «que sea técnicamente factible»; moviéndose más en el plano de la recomendación y del posibilismo, que en el de la obligación.

Detengámonos un momento en comprender el proceso de inferencia de la IA. La capacidad de inferencia es el proceso en el que un modelo de IA entrenado aplica lo que ha aprendido de nuevos datos para hacer predicciones, contenidos, recomendaciones o tomar decisiones. Las técnicas que permiten la inferencia al construir un sistema de IA incluyen estrategias de aprendizaje automático que aprenden de los datos cómo alcanzar determinados objetivos y estrategias basadas en la lógica y el conocimiento que infieren a partir de conocimientos codificados o de una representación simbólica de la tarea que debe resolverse.

Se ha alcanzado actualmente, entre la doctrina científica, cierto consenso sobre la urgencia de un marco regulador de los procesos de diseño, evaluación y control de estos nuevos sistemas en el sector público, al menos cuando estos puedan afectar a los derechos fundamentales de las personas, que permita adoptar medidas que eviten la opacidad con la que operan, evitando que existan cajas negras administrativas (Cerrillo i Martínez, 2019 y Velasco Rico, 2024) o sesgos que den lugar a resoluciones administrativas injustas o discriminatorias, o que pudiesen vulnerar derechos fundamentales de la ciudadanía.

Algunos autores plantean que el empleo de algoritmos que predeterminen la toma de decisiones administrativas directamente han de ser considerados reglamentos y, en consecuencia, aplicarles las mismas garantías tradicionales que nuestro ordenamiento jurídico prevé para las normas reglamentarias (Boix Palop, 2020); básicamente las que tiene que ver con la necesidad de una planificación normativa, una evaluación ex-ante, que identifique con precisión lo que se pretende conseguir y las medidas adecuadas para ello; una evaluación ex-post, al objeto de revisar si se han conseguido los objetivos planteados o no; así como de medidas de fomento de la participación ciudadana en el proceso.



En el ámbito del derecho de la Unión Europea, la solución al empleo de algoritmos par la toma de decisiones administrativas se centró inicialmente en dictar normas en materia de protección de datos, lo que era absolutamente insuficiente.

Con fecha 13 de marzo de 2024 se aprobó la Resolución legislativa del Parlamento Europeo, sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo, por el que se establecen normas armonizadas en materia de inteligencia artificial y se modifican determinados reglamentos de la Unión (Comisión Europea, 2021). La elección del Reglamento como instrumento jurídico, se debe a la necesidad de aplicar uniformemente las nuevas normas, tales como la definición de IA, la prohibición de determinadas prácticas perjudiciales que la IA permitiría y la clasificación de determinados sistemas de IA. Puesto que, de conformidad con el art. 288 del TFUE<sup>4</sup>, los reglamentos son directamente aplicables, la elección de este instrumento reducirá la fragmentación jurídica y facilitará el desarrollo de un mercado único de sistemas de IA legales, seguros y fiables (Presno, 2023).

El Reglamento, que espera publicarse en el DOUE en julio de 2024 y entrará en vigor veinte días después de su publicación, será de plena aplicación veinticuatro meses después de ello, con excepción de: las prohibiciones de determinadas prácticas (que se aplicarán seis meses después de la fecha de entrada en vigor); los códigos de buenas prácticas (nueve meses después); las normas sobre la IA de uso general, incluida su gobernanza (doce meses después), y las obligaciones para los sistemas de alto riesgo (treinta y seis meses después).

El ya próximo Reglamento europeo adopta un enfoque basado en la gestión de riesgos. Categoriza los riesgos en cuatro niveles, que van desde el riesgo mínimo, riesgo limitado, riesgo alto y riesgo inaceptable.

Entre los riesgos inaceptables, se encuentran, entre otros, los sistemas de IA que se consideren una amenaza para la seguridad, los derechos fundamentales o los valores de la UE, como podrían ser los sistemas que manipulan el comportamiento de las personas empleando técnicas subliminales o deliberadamente manipuladoras o engañosas; los sistemas que exploten alguna de las vulnerabilidades de una persona o grupo específico con el objetivo de alterar su comportamiento; los sistemas de evaluación o clasificación de personas físicas o grupos atendiendo a su comportamiento social o características personales o de personalidad, de forma que su puntuación social inferida provoque un trato perjudicial o desfavorable (el Sistema de Crédito Social de China es un ejemplo en el uso de este tipo de sistemas prohibidos en el reglamento de la UE); los sistemas de IA para realizar evaluaciones de riesgos de personas físicas con el fin de evaluar o predecir la probabilidad de que una persona física cometa una infracción penal; sistemas de IA para inferir las emociones de una persona física en los lugares de trabajo y en los centros educativos, excepto cuando el sistema de IA esté destinado a ser instalado o introducido en el mercado por motivos médicos o de seguridad; los sistemas de categorización biométrica que clasifiquen individualmente a las personas físicas sobre la base de sus datos biométricos para deducir o inferir su raza, opiniones políticas, afiliación sindical, convicciones religiosas o filosóficas, vida sexual u orientación sexual; el uso de sistemas de identificación biométrica remota «en tiempo real» en espacios de acceso público con fines de aplicación de la ley, salvo en determinados supuestos; etc.

Califica de riesgo alto a los sistemas cuya salida puede impactar significativamente en la seguridad o derechos fundamentales, como los de identificación biométrica en tiempo real, los de protección de infraestructuras críticas, de selección y promoción de personal, los usados por los Cuerpos y Fuerzas de Seguridad del Estado o de la Administración de justicia). Estos sistemas están sujetos a mayores obligaciones de transparencia y deben realizar una evaluación del impacto en los derechos fundamentales antes de ser puestos en explotación.

Con este enfoque las Instituciones de la UE declaran favorecer el desarrollo y adopción de una «Inteligencia Artificial fiable y segura, centrada en el ser humano»; si bien no faltan análisis críticos sobre las insuficiencias de esta regulación y la influencia de las grandes compañías tecnológicas para decantar hacia sus intereses («captura regulatoria») importantes aspectos de esta normativa.

### **3.2. La posible afección a los derechos fundamentales de la ciudadanía por errores, sesgos y discriminaciones algorítmicas**

El empleo de «Gervasio», como el de cualquier sistema de IA, no está exento de la posibilidad de que se generen sesgos y discriminaciones, con impactos que, en algunos casos, pueden llegar a ser muy

<sup>4</sup> Tratado de Funcionamiento de la Unión Europea.

relevantes sobre los derechos de los ciudadanos. A estos efectos, conviene recordar que la Constitución Española de 1978 establece en su art. 18.4 que «La ley limitará el uso de la informática para garantizar el honor y la intimidad personal y familiar de los ciudadanos y el pleno ejercicio de sus derechos». Por tanto, el poder legislativo dispone del mandato derivado de nuestra norma fundamental para establecer los límites a los sistemas de IA que se juzguen oportunos para proteger los derechos de los ciudadanos.

Se habla de «discriminación algorítmica» para referirse a la que «se produce cuando los sistemas automatizados contribuyen a un trato diferente injustificado» o, en su caso, un «impacto que desfavorece a las personas» en función de circunstancias específicas especialmente sospechosas de discriminación (raza, color, sexo, religión, edad, origen nacional, etc.) de modo contrario a Derecho.

El concepto de sesgo algorítmico acuñado por el Grupo Independiente de Expertos de Alto nivel sobre Inteligencia Artificial (AI-HLEG, 2019) nombrados por la Comisión Europea, es el de «una inclinación que favorece o perjudica a una persona, objeto o posición». Se trata de «una anomalía en la salida de los sistemas de IA, debido a los prejuicios o suposiciones erróneas realizadas durante el proceso de desarrollo del sistema o prejuicios en los datos de entrenamiento, por lo que los resultados del sistema de IA no pueden generalizarse ampliamente».

El Instituto Nacional de Estándares y Tecnología de los Estados Unidos (National Institute of Standards and Technology, NIST), ha publicado la versión definitiva de su documento sobre la definición de unos modelos para identificar y gestionar los sesgos en la Inteligencia Artificial, *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence* (Schwartz *et al.*, 2022).

El NIST clasifica los diferentes tipos de sesgos de la IA en tres categorías dominantes (Fernández Hernández, 2022):

- Los sesgos sistémicos, que son el resultado de los procedimientos y las prácticas de determinadas instituciones que operan de manera que dan lugar a que ciertos grupos sociales se vean favorecidos y otros se vean perjudicados o devaluados. Esto no tiene por qué ser el resultado de ningún prejuicio o discriminación consciente, sino de que la mayoría siga las reglas o normas existentes. El racismo y el sexismo institucionales son los ejemplos más comunes.
- Los sesgos estadísticos y computacionales, que se derivan de los errores que se producen cuando la muestra no es representativa de la población. Estos sesgos surgen de un error sistemático y no aleatorio y pueden producirse en ausencia de prejuicios, parcialidad o intención discriminatoria. En sistemas de IA, estos sesgos están presentes en los conjuntos de datos y en los procesos algorítmicos utilizados en el desarrollo de aplicaciones de IA, y a menudo surgen cuando los algoritmos se entrenan con un tipo de datos y no pueden extrapolar más allá de esos datos. El error puede deberse a la heterogeneidad de los datos heterogéneos, representación de datos complejos en representaciones matemáticas más simples, datos erróneos y a sesgos algorítmicos como el sobreajuste y el infraajuste, el tratamiento de los valores atípicos y los factores de limpieza e imputación de datos, y factores de limpieza e imputación de datos. Datos de baja calidad llevan a resultados de baja calidad (Cotino Hueso, 2023).
- Los sesgos humanos, reflejan errores sistemáticos del pensamiento humano basados en un número limitado de principios heurísticos y la predicción de valores a operaciones de juicio más simples. Estos sesgos son a menudo implícitos y tienden a relacionarse con la forma en que un individuo o grupo percibe la información (como la salida automatizada de la IA) para tomar una decisión o rellenar la información que falta o se desconoce. información desconocida. Estos sesgos están omnipresentes en los procesos de toma de decisiones institucionales, grupales e individuales a lo largo del ciclo de vida de la IA, y en el uso de las aplicaciones de IA una vez desplegadas. Así, se dan sesgos de anclaje, de disponibilidad, de confirmación, entre muchos otros.

Los supuestos documentados de discriminación algorítmica en el uso público y privado de la IA en todos los países son numerosos. Así, por ejemplo, se han detectado en sistemas de identificación policial o predictivos de riesgos de reincidencia delictivas, en sistemas de asignación de ayudas públicas y beneficios sociales, en sistemas de IA aplicados al ámbito educativo para predecir abandono escolar o evaluar a docentes, en el ámbito del empleo, etc. En España, entre otros supuestos, el Gobierno estatal creó la aplicación BOSCO, cedida a las compañías eléctricas para verificar los requisitos de los solicitantes para la atribución del bono social eléctrico. Su funcionamiento ha generado sospechas de errores y sesgo de discriminación, que no se han podido probar plenamente por la reiterada falta de transparencia sobre el mismo. La reclamación ante los Tribunales realizada desde 2018 por la organización Civio para que se revele el código fuente de BOSCO ha sido desestimada, tanto en primera instancia, como por la Audiencia Nacional.

En muchos casos la discriminación o sesgos algorítmicos por un sistema de IA público se deberán a decisiones adoptadas en el desarrollo, implantación y uso del sistema, probablemente no formalizadas; pero, aunque no adquieran forma jurídica o formen parte del procedimiento, debe ser reparada, al amparo del incidente discriminatorio previsto en el art. 25. 3.º de la Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación.

### 3.3. Para la protección de los derechos de autor y de los datos personales

El Reglamento Europeo especifica que no se podrá entrenar a los sistemas de IA de propósito general con materiales sujetos a copyright siempre y cuando sus autores hayan expresado de forma específica que estos no pueden ser utilizados para el entrenamiento de sistemas de IA. Pero, existen muchas dudas sobre la aplicación práctica de esta prohibición.

En cuanto a la protección de datos personales, pese a que en la UE ya existe una normativa extensa sobre esta cuestión<sup>5</sup>, no son pocas las voces autorizadas que consideran que la IA puede acentuar los problemas actuales sobre filtración de datos privados, apropiación indebida de datos, tergiversación y violación de la intimidad. Además, existe la preocupación de que, con una pequeña cantidad de datos y una interacción suficientemente prolongada con los usuarios, los sistemas de aprendizaje profundo sean capaces de intuir otros datos de estos y registrarlos sin su consentimiento.

### 3.4. La necesaria sujeción de la IA a principios éticos

Los desafíos apuntados en los apartados anteriores están fuertemente relacionados con la necesidad de aplicar principios éticos a los sistemas de IA.

Son numerosos los pronunciamientos de Instituciones Internacionales acerca de que la IA debe estar presidida por la ética. La Unión Europea incluso, ha adoptado la estrategia de intentar situarse a la vanguardia del mundo en el desarrollo de una IA ética y confiable (Comisión Europea, 2020), como marca diferenciadora frente a las grandes potencias en esta tecnología, como EE. UU. y China (Criado, 2021; Cotino Hueso, 2019), que no otorgan especial relevancia a esta cuestión<sup>6</sup>.

Como el propio Reglamento europeo declara, se pretende «el establecimiento de un marco jurídico uniforme, en particular para el desarrollo, la introducción en el mercado, la puesta en servicio y la utilización de sistemas de inteligencia artificial (“sistemas de IA”) en la Unión, de conformidad con los valores de la Unión, a fin de promover la adopción de una inteligencia artificial (IA) centrada en el ser humano y fiable, garantizando al mismo tiempo un elevado nivel de protección de la salud, la seguridad y los derechos fundamentales consagrados en la Carta de los Derechos Fundamentales de la Unión Europea (la “Carta”), en particular la democracia, el Estado de Derecho y la protección del medio ambiente, frente a los efectos perjudiciales de los sistemas de IA en la Unión, así como brindar apoyo a la innovación». Normas que, también señala, también deben tener en cuenta la Declaración Europea sobre los Derechos y Principios Digitales para la Década Digital y las Directrices éticas para una IA fiable del Grupo de Expertos de Alto Nivel sobre la Inteligencia Artificial.

En las Directrices éticas para una IA fiable (AI-HLEG, 2019), el Grupo Independiente de Expertos de Alto Nivel sobre IA creado por la Comisión desarrolló siete principios éticos no vinculantes para la IA, que tienen por objeto contribuir a garantizar la fiabilidad y el fundamento ético de la IA.

Los siete principios son:

- *Acción y supervisión humanas*: Los sistemas de IA deben desarrollarse y utilizarse como herramienta al servicio de las personas, respetando la dignidad humana y la autonomía personal, y funcionar de manera que puedan ser controlados y vigilados adecuadamente por seres humanos.
- *Solidez técnica y seguridad*: Los sistemas de IA se deben desarrollar y utilizar de manera que sean sólidos en caso de problemas; resilientes frente a los intentos de alterar su uso por terceros para fines ilícitos; buscando minimizar los daños no deseados.

<sup>5</sup> El derecho fundamental a la protección de los datos personales está garantizado, en particular, por los Reglamentos (UE) 2016/679 y (UE) 2018/1725 del Parlamento Europeo y del Consejo y la Directiva (UE) 2016/680 del Parlamento Europeo y del Consejo. Además, la Directiva 2002/58/CE del Parlamento Europeo y del Consejo protege la vida privada y la confidencialidad de las comunicaciones, incluso estableciendo condiciones para el almacenamiento de datos personales y no personales en los equipos terminales, y el acceso desde estos.

<sup>6</sup> Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas [2020/2012(INL)].

- *Gestión de la privacidad y de los datos*: Los sistemas de IA se deben desarrollar y utilizar de conformidad con las normas en materia de protección de la intimidad y de los datos. Asimismo, deben tratar datos que cumplan normas estrictas de calidad e integridad.
- *Transparencia*: los sistemas de IA se deben desarrollar y utilizar de un modo que permita una trazabilidad adecuada, y que, al mismo tiempo, haga que las personas sean conscientes de que se comunican o interactúan con un sistema de IA, e informe debidamente a los responsables del despliegue acerca de las capacidades y limitaciones de dicho sistema de IA y a las personas afectadas acerca de sus derechos.
- *Diversidad, no discriminación y equidad*: Los sistemas de IA se deben desarrollar y utilizar de un modo inclusivo, promoviendo la igualdad de acceso, la igualdad de género y la diversidad cultural, al tiempo que evitan los efectos discriminatorios y los sesgos injustos o prohibidos por el Derecho nacional o de la Unión.
- *Bienestar social y ambiental*: Los sistemas de IA se deben desarrollar y utilizar de manera sostenible y respetuosa con el medio ambiente, así como en beneficio de todos los seres humanos, al tiempo que se supervisan y evalúan los efectos a largo plazo en las personas, la sociedad y la democracia.
- *Rendición de cuentas*: Implica la necesidad de entender y pedir cuentas. Debe poderse responder a la pregunta «¿cómo funciona?». Y también en el sentido ético de responsabilidad: «¿quién es responsable de la forma en que funciona?».

Se trata de principios cuya implementación y auditoría no parece sencilla.

Así, por ejemplo, respecto al principio de supervisión humana, ¿La organización Open AI, desarrolladora de Chat GPT, va a explicarnos el funcionamiento de su sistema algorítmico cuando consideremos integrarla en alguna de las aplicaciones del sector público? ¿Y, respecto a la gestión de la privacidad y de los datos, pese a que ya nos hemos lanzado a integrar la IA, disponemos en las Administraciones públicas de estrategias de preparación de datos (enfocadas al acceso a datos de calidad y a la maximización de su uso en las operaciones) y de infraestructuras de datos necesarias que ofrezcan confianza en materia de seguridad? El Estudio de Forrester Consulting de enero de 2024<sup>7</sup>, encargado por Salesforce, sobre organizaciones del sector empresarial señala lo contrario; por lo que es previsible que en el sector público ocurra lo mismo.

Parece que es urgente desarrollar sistemas y herramientas de IA que nos ayuden a controlar a la propia IA.

#### 4. EL ROL A DESEMPEÑAR POR LA POLÍTICA Y LA ADMINISTRACIÓN PÚBLICA ANTE LOS POSIBLES IMPACTOS DESFAVORABLES DE LA IA SOBRE LA SOCIEDAD

Sin perjuicio de que las autoridades nacionales deben jugar un papel de supervisión de la aplicación del próximo reglamento europeo sobre la IA<sup>8</sup>; las Administraciones como desarrolladoras, promotoras o usuarias de sistemas de IA, van a tener que adoptar un abanico de medidas preventivas y de evaluación; así como de reingeniería de los procesos de control, que abarque a los proveedores relacionados con la contratación de IA y servicios o productos generados por IA.

Desde la óptica interna, parece evidente que el diseño de soluciones innovadoras para afrontar o mitigar los previsible impactos que la IA va a provocar sobre las políticas públicas van a exigir profundas transformaciones que afectarán a la estructura organizativa y a los perfiles políticos, directivos y de los empleados públicos que las conforman.

Resulta imposible, por razones obvias, hacer referencia a la repercusión que los sistemas de IA avanzados podrían tener sobre la totalidad de las políticas públicas; así que tomemos como ejemplo su previsible impacto en algunas de ellas.

##### 4.1. Sobre el empleo

Como señala Harari (2018), los humanos tienen dos tipos de capacidades: la física y la cognitiva. En el pasado las máquinas competían con los humanos principalmente por las capacidades físicas en bruto.

<sup>7</sup> Forrester Consulting (2024). The Journey to AI-Powered CRM. Salesforce. <https://staging.wisdominterface.com/wp-content/uploads/2024/06/Salesforce-AI-CRM-TLP.pdf>

<sup>8</sup> En España, se ha creado a tal efecto la Agencia Española de Supervisión de la Inteligencia Artificial («AESIA»).



Perdida la batalla por los humanos en torno a las competencias físicas, la IA está empezando a superarlos también en capacidades cognitivas que antes solo tenían estos, como las de aprender, analizar, comunicar y, por si fuera poco, también en ... comprender las emociones humanas y en creatividad.

Más allá de las nuevas imágenes y canciones que ya elabora la IA interpretando nuestras peticiones, la batalla, en diciembre de 2017, por convertirse en el mejor motor de ajedrez del mundo, entre los programas *Stockfish* y *AlphaZero*, puede ser un buen ejemplo del poder innovador y creativo de un sistema de aprendizaje profundo.

*Stockfish* era el programa campeón mundial de ajedrez en 2016. Era un motor basado en métodos tradicionales de búsqueda y evaluación de posiciones, con acceso a siglos de experiencia humana acumulada en ajedrez y décadas de experiencia como ordenador. Podía calcular 70 millones de posiciones en el tablero por segundo. *Alpha Zero*, en cambio, era un programa nuevo, desarrollado por *DeepMind*, sirviéndose de algoritmos de aprendizaje automático profundo, que aprendió a jugar al ajedrez desde cero, entrenando consigo mismo durante solo nueve horas antes del enfrentamiento. *AlphaZero* solo realizaba 80 mil cálculos por segundo y nunca fue entrenado por humanos sobre estrategias de ajedrez.

Jugaron 100 partidas, de las que *AlphaZero* ganó 28 con piezas blancas y 25 con negras. El resto (47) terminaron en tablas. Mostró un estilo de juego creativo, innovador y agresivo, a menudo sacrificando piezas para obtener ventajas estratégicas a largo plazo. La victoria de *AlphaZero* demostró la eficacia del enfoque de aprendizaje profundo para resolver problemas complejos sin necesidad de conocimiento humano previo. En solo unas horas *Alpha Zero* pasó de la ignorancia absoluta en ajedrez a Gran Maestro sin ayuda humana.

Aunque no participamos de las elucubraciones conspiranoicas que vaticinan la extinción de la civilización humana por culpa de los robots inteligentes, la realidad es que nadie sabe con certeza cómo serán los efectos de la IA sobre el empleo en el futuro. Como vemos a continuación, las estimaciones de algunos organismos dedicados a la prospectiva varían considerablemente:

- Según el World Economic Forum (WEF, 2023), *The Future of Jobs Report*, el auge de la IA y la transición verde transformarán la forma en que trabajamos. Predicen que el 42% de las tareas empresariales se automatizarán para 2027. Se espera que casi el 75% de las empresas encuestadas adopten la inteligencia artificial, un impulsor clave del posible desplazamiento algorítmico, y se espera que genere una alta rotación: el 50% de las organizaciones espera que genere crecimiento de empleo y el 25% espera que produzca pérdidas de empleo.
- Para el McKinsey Global Institute (2017) sus escenarios sugieren que para 2030, entre 75 millones y 375 millones de trabajadores (del 3 al 14% de la fuerza laboral global) necesitarán cambiar de categoría ocupacional.
- Gartner (2024) señala como una de las 9 tendencias que definirán el futuro del trabajo, la de que la IA crea oportunidades para la fuerza laboral, no las elimina. La implementación de la IA generativa no solo creará la necesidad de nuevos puestos, sino que permitirá a los empleadores reducir el tiempo para que un empleado sea productivo con las nuevas tecnologías y especializaciones, lo que aliviará la necesidad de contratar talento ya capacitado en estas áreas.
- La OCDE en 2019 estimaba que el 14% de los empleos de los países inscritos en dicha organización estaban en alto riesgo de automatización; y que el 32% de los empleos podrían experimentar cambios significativos en sus tareas debido a la automatización y la IA.

Pese a la variabilidad de las estimaciones, resulta evidente que la IA va a ocasionar un impacto muy relevante sobre el empleo, y, sin embargo, no aparece esta preocupación en los programas electorales de los partidos políticos. Algunos líderes políticos andan muy preocupados culpando de la pérdida de empleo a la inmigración, cuando es muchísimo más probable que se deba a las profundas transformaciones que va a producir la IA. ¿Querrán que los efectos desfavorables los solucionen los miembros de los Consejos de Administración de las compañías tecnológicas o los programadores; pese a que nos lo ha votado nadie?

En cualquier caso, como señala Huergo Lora (2023), es una evidencia empírica que ningún avance tecnológico se ha detenido por la destrucción de puestos de trabajo.

## 4.2. Sobre la movilidad

Algunos estudios de prospectiva sitúan el desenvolvimiento del vehículo autónomo, aquéllos capaces de detectar su entorno y navegar sin intervención humana, en un horizonte temporal no muy lejano. La hoja de ruta para la conducción autónoma publicada por la Plataforma Tecnológica Europea, que reúne a expertos



de la industria, la investigación y las instituciones públicas, prevé en su versión de 2017 que el desarrollo de vehículos completamente autónomos se producirá durante la década de 2030 (Barrio Andrés, 2019).

Según la Organización Mundial de la Salud (OMS) se producen en el Planeta 1,3 millones de muertes por accidentes de tráfico en el mundo, a lo que habría que añadir la estimación de que entre 20 y 50 millones de personas sufren lesiones o discapacidades graves al año por este tipo de accidentes.

Teniendo en cuenta las causas más frecuentes de los accidentes de tráfico se deben a errores humanos (más del 90 %), parece claro que los algoritmos que asuman la navegación autónoma de los vehículos no infringirán los límites de velocidad, no conducirán bajo los efectos del alcohol o drogas, no se distraerán mientras conducen, ni se quedarán dormidos al volante; por lo que podrían evitar más de 1 millón de muertes al año, y mucho dolor, secuelas y gasto sanitario derivado de lesiones graves; por lo que impedir la automatización en el campo del transporte no parece buena idea.

Ahora bien, el vehículo autónomo también producirá, gradualmente claro, la desaparición de conductores y de policías de tráfico.

Por ejemplo, en España, según datos del INE a 2020, existen 116 mil conductores de vehículos de transporte regular de viajeros. ¿Cuándo la implantación del vehículo autónomo sea una realidad, qué pasará con el empleo de estas personas, ¿Cómo los poderes públicos afrontarán sus demandas?, ¿Podrán las Administraciones poner en marcha políticas para cualificarlos en otras profesiones?, ¿Y mientras tanto se forman en, por ejemplo, operador/a de drones o en diseñadores de mundos virtuales, habrá suficientes subsidios para garantizarles a todos sus necesidades básicas?

### 4.3. Para la sostenibilidad medioambiental

La inteligencia artificial (IA) tiene un impacto significativo en la sostenibilidad medioambiental del planeta, con efectos tanto positivos como negativos.

La influencia positiva de la IA viene de la mano, por ejemplo, de sus capacidades para disponer de Redes eléctricas inteligentes (optimizando la eficiencia en el funcionamiento de las redes de generación y distribución de electricidad, o efectuando predicciones de la demanda energética con precisión para introducir mejoras en su gestión); de su potencial para el monitoreo y análisis de los ecosistemas para detectar cambios ambientales, predecir desastres naturales; de su capacidad para mejorar la eficiencia y reducir el impacto medioambiental en todos los sectores económicos; de su uso para mejorar la eficiencia en las edificaciones, etc.

Centrándonos en los negativos. Podemos advertir los siguientes:

#### 4.3.1. Consumo de energía de los Centros de Datos

Tanto los procesos de capacitación de la IA como de inferencia a gran escala requieren sistemas informáticos centralizados y de alto rendimiento optimizados para cargas de trabajo de IA alojadas en centros de datos. Debido a la inmensa escala de implementación del modelo actual, la mayor parte de todo el cálculo de la IA ahora se utiliza para inferencia, aunque una sola ejecución de entrenamiento requiere mucha más computación que una sola inferencia.

Por ejemplo, las aplicaciones de IA ampliamente utilizadas, como la búsqueda en Internet, el reconocimiento de voz y la traducción de idiomas, requieren una infraestructura informática a gran escala para atender a miles de millones de usuarios. Los centros de datos de las grandes empresas tecnológicas pueden consumir tanta energía como la de miles de hogares o incluso pequeñas ciudades.

Un estudio de la Universidad de Massachusetts Amherst (Strubell *et al.*, 2019), señaló que un modelo de IA de Procesamiento de Lenguaje Natural, como el modelo GPT-2, consume aproximadamente 85,34 GJ (gigajulios) de energía durante su entrenamiento, lo que produce unas emisiones de alrededor de 626,155 kg de CO<sub>2</sub>.

En definitiva, las aplicaciones de IA avanzada, especialmente aquellas que involucran el entrenamiento de modelos de aprendizaje profundo, son particularmente intensivas en términos de recursos computacionales y, por ende, de energía, lo cual contribuye a incrementar las emisiones de carbono si la fuente de energía no es renovable.

#### 4.3.2. Residuos electrónicos

La rápida evolución de la IA puede llevar a una mayor obsolescencia de hardware, resultando en un incremento de residuos electrónicos que deben ser gestionados adecuadamente para minimizar el impacto ambiental.

### 4.3.3. Impacto del ciclo de vida de los dispositivos

La producción de dispositivos electrónicos y la disposición de los mismos tienen impactos ambientales significativos, desde la extracción de materiales hasta el desecho final.

Este impacto de la IA en la sostenibilidad medioambiental va a depender del grado en que se implementen políticas y regulaciones que promuevan el uso de energías renovables y de eficiencia energética en los centros de datos; de las innovaciones en el diseño de hardware y algoritmos más eficientes que requieran menos recursos computacionales para ser eficaces; así como de que las empresas y organizaciones adopten prácticas de responsabilidad social corporativa destinadas a reducir la huella de carbono de sus operaciones basadas en IA.

## 5. ALGUNAS REFLEXIONES PARA CONCLUIR

En los albores de la cuarta revolución industrial, en la que el papel de la Inteligencia Artificial está llamado a ser determinante, es innegable que el sector público puede y debe promover el uso de la misma para optimizar la creación de valor público y el bienestar común; pero en este objetivo debe ponderar la opacidad y los riesgos que presentan ya los sistemas de IA avanzados, y adoptar las medidas que considere necesarias para proteger los derechos fundamentales de las personas y los valores democráticos recogidos tanto en nuestra Constitución, como en el ordenamiento de la Unión Europea, reforzando la legitimidad de sus decisiones cuando en su predeterminación se empleen estas herramientas.

La rápida generalización por todas las Administraciones públicas del uso de sistemas de IA, espoleada por la vertiginosa carrera de avances en esta materia y sus enormes potencialidades para mejorar el desempeño del sector público y la renovación de su cartera de servicios, hace que los sistemas algorítmicos se perciban ya como el principal elemento configurador del «sistema operativo» de estas, de manera que la «Gobernanza Pública Algorítmica» es ya una realidad.

Pese a ello, existe consenso doctrinal en la necesidad de que el sector público se dote de un marco normativo básico sobre la IA que supere las insuficiencias detectadas en la normativa vigente; toda vez que contagiadas por la aceleración de los constantes avances, el sector público se ha lanzado a promover el empleo de estos sistemas sin que la normativa reguladora del uso de la IA por la Administración se haya actualizado a la misma velocidad. Asimismo, la aplicación de principios éticos, desde la fase de diseño, de los sistemas de IA es una estrategia europea que parece acertada, pese a plantear interrogantes sobre si su aplicación puede constituir un freno a su desarrollo, frente a otros ámbitos territoriales en los que la cuestión ética no aparece como requerimiento; o el de su efectiva aplicación y control. En cualquier caso, parece urgente la necesidad de construir soluciones de IA que ayuden a controlar a la propia IA.

Más allá del rol que las Administraciones públicas desempeñen en el fomento de estas herramientas y en la dinamización del mercado; y de su papel regulador y de supervisión; las Administraciones públicas van a tener que desempeñar un importante papel en la resolución de problemáticas de diversa índole generada por la aplicación de la IA en la sociedad, enfrentándose a desafíos inéditos hasta ahora; y para ello será preciso también que se produzcan transformaciones que afectarán a su estructura organizativa y a los perfiles políticos, directivos y de los empleados públicos que las conforman.

## REFERENCIAS BIBLIOGRÁFICAS

- Barrio Andrés, M. (2019). El coche autónomo acelera. *El País*.
- Boix Palop, A. (2020). Los algoritmos son reglamentos: la necesidad de extender las garantías propias de las normas reglamentarias a los programas empleados por la administración para la adopción de decisiones. *Revista de Derecho Público: Teoría y Método*, 1(1), 223-270. [https://doi.org/10.37417/RPD/vol\\_1\\_2020\\_33](https://doi.org/10.37417/RPD/vol_1_2020_33)
- Cerrillo i Martínez, A. (2019). El impacto de la inteligencia artificial en el derecho administrativo, ¿nuevos conceptos para nuevas realidades técnicas? *Revista General de Derecho Administrativo*, (50). [https://www.iustel.com/lv2/revistas/detalle\\_revista.asp?id\\_noticia=421172](https://www.iustel.com/lv2/revistas/detalle_revista.asp?id_noticia=421172)
- COMISIÓN EUROPEA (2001). *Comunicación de la Comisión, de 25 de julio de 2001, «La gobernanza europea – Un Libro Blanco»* [COM (2001) 428 final – Diario Oficial C 287 de 12-10-2001]. <https://eur-lex.europa.eu/ES/legal-content/summary/white-paper-on-governance.html>
- COMISIÓN EUROPEA (2020). *Libro Blanco sobre la inteligencia artificial. Un enfoque europeo orientado a la excelencia y la confianza*. Oficina de Publicaciones de la Unión Europea. <https://op.europa.eu/es/publication-detail/-/publication/ac957f13-53c6-11ea-aece-01aa75ed71a1>

- COMISIÓN EUROPEA (2021). *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/2021/206 final]. <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:52021PC0206>
- Cotino Hueso, L. (2019). Ética en el diseño para el desarrollo de una inteligencia artificial, robótica y big data confiables y su utilidad desde el derecho. *Revista Catalana de Dret Públic*, (58), 29-48. <https://doi.org/10.2436/rcdp.158.2019.3303>
- Cotino Hueso, L. (2023). Discriminación, sesgos e igualdad de la inteligencia artificial en el sector público. En E. Gamero Casado y F. L. Pérez Guerrero (coords.), *Inteligencia artificial y sector público: retos, límites y medios* (pp. 257-351). Tirant lo Blanch. Disponible en: <https://www.uv.es/cotino/publicaciones/publicadoCOTINOsosgos.pdf>
- Criado, J. I. (2021). Inteligencia Artificial (y Administración pública). *Eunomia. Revista en Cultura de la Legalidad*, (20), 348-372. <https://doi.org/10.20318/eunomia.2021.6097>
- Fernández Hernández, C. (2022). Modelos para identificar y gestionar los sesgos en la Inteligencia Artificial. *Diario La Ley*, (60). [https://diariolaley.laleynext.es/Content/DocumentoRelacionado.aspx?params=H4sIAAAAAAAAAEAC2NQUVvEQAYFf425CDJdsauHXLrrTUS0iNd0JrQD7WTNZOr23zu6G3iER768911Yt57Phl5SLksMore8nJQzQd6SpG3BXguD0ZDR3ex9U7UD8lZoPorH-9b9ubhyTwPuQDSwdhs6MDGa3zlj\\_gAeZKfV1rjSBYlDaSX2BgCPn-5O3seNe0TrKy5AvgZR07GMMVxeqmyC5-Z1E9vNDLW8rJURu4on87XS1fM6vdg6ePfg5\\_rPpLxgWZO4dr7C\\_9RzHD4AAAAWKE](https://diariolaley.laleynext.es/Content/DocumentoRelacionado.aspx?params=H4sIAAAAAAAAAEAC2NQUVvEQAYFf425CDJdsauHXLrrTUS0iNd0JrQD7WTNZOr23zu6G3iER768911Yt57Phl5SLksMore8nJQzQd6SpG3BXguD0ZDR3ex9U7UD8lZoPorH-9b9ubhyTwPuQDSwdhs6MDGa3zlj_gAeZKfV1rjSBYlDaSX2BgCPn-5O3seNe0TrKy5AvgZR07GMMVxeqmyC5-Z1E9vNDLW8rJURu4on87XS1fM6vdg6ePfg5_rPpLxgWZO4dr7C_9RzHD4AAAAWKE)
- GARTNER (2024). *Las 9 tendencias que definirán el futuro del trabajo en 2024*. <https://www.gartner.es/es/articulos/las-9-tendencias-que-definiran-el-futuro-del-trabajo-en-2024>
- GRUPO INDEPENDIENTE DE EXPERTOS DE ALTO NIVEL SOBRE INTELIGENCIA ARTIFICIAL – AI-HLEG (2019). *Directrices éticas para una IA fiable*. Oficina de Publicaciones de la Unión Europea. <https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>
- Harari, Y. N. (2018). *21 lecciones para el siglo XXI*. Debate.
- Huergo Lora, A. (2023). Inteligencia artificial: una aproximación jurídica no catastrofista. *Revista Española de Control Externo*, (74-75), 110-129. [https://recex.tcu.es/export/sites/nuevo-recex/galleries/pdf/R74\\_ART-7-AHL.pdf](https://recex.tcu.es/export/sites/nuevo-recex/galleries/pdf/R74_ART-7-AHL.pdf)
- MCKINSEY GLOBAL INSTITUTE (2017). *Jobs lost, Jobs gained: Workforce transitions in a time of automation*. <https://www.mckinsey.com/~media/BAB489A30B724BECB5DEDC41E9BB9FAC.ashx>
- Prats i Catalá, J. (2005). *De la burocracia al management, del management a la gobernanza*. INAP.
- Presno Linera, M. A. (2023). *La propuesta de «Ley de Inteligencia Artificial» europea*. *Revista de Las Cortes Generales*, (116), 81-133. <https://doi.org/10.33426/rcg/2023/116/1775>
- Reinsel, D., Gantz, J. y Rydning, J. (2017). *Data Age 2025: The Evolution of Data to Life-Critical* [IDC White Paper]. Seagate and International Data Computing. <https://www.seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf>
- Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A. y Hall, P. (2022). *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*. National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.SP.1270>
- Strubell, E., Ganesh, A. y McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. *Annual Meeting of the Association for Computational Linguistics*, (5). <https://arxiv.org/pdf/1906.02243>
- Velasco Rico, C. I. (2024). *Marco regulatorio de los sistemas algorítmicos y de Inteligencia Artificial: el papel de la Administración*. Asociación Española de Profesores de Derecho Administrativo.
- WORLD ECONOMIC FORUM (2023). *The Future of Jobs Report 2023*. <https://www.weforum.org/publications/the-future-of-jobs-report-2023/digest/>