

EL GOBIERNO DE LOS SISTEMAS DECISORIOS HÍBRIDOS EMPLEADOS EN ACTUACIONES ADMINISTRATIVAS¹.

Juanita Pedraza Córdoba

RESUMEN

Esta comunicación versa sobre el gobierno de los sistemas decisorios híbridos, en los que las personas interactúan con sistemas, con el fin de adoptar decisiones administrativas. El gobierno se define en un doble plano: normativo y organizativo.

ABSTRACT

This working paper is about the governance of hybrid decision-making systems, in which people interact with systems in order to make administrative decisions. The government is defined on a double level: regulatory and organizational.

PALABRAS CLAVES

Reserva de humanidad, human in the loop, human out the loop, human on the loop

KEY WORDS

Reserved for humans, human in the loop, human out the loop, human on the loop

SUMARIO: 1.INTRODUCCIÓN, 2. CONTEXTUALIZACIÓN Y DELIMITACIÓN CONCEPTUAL, 3. CONDICIONES DE COLABORACIÓN EN EL PLANO NORMATIVO, 4. CONDICIONES DE COLABORACIÓN EN EL PLANO ORGANIZATIVO, A. Elección de enfoque y planificación, B. Delimitación de la función correctiva, 5. REFERENCIAS

1. INTRODUCCIÓN

El proceso de transformación digital de la administración española halla su origen en el art. treinta.dos de la Ley de 17 de julio de 1958, del procedimiento administrativo y discurre al amparo de varias normas (Ley 30/1992, Ley 11/2007), hasta llegar a la formulación actual de la administración como plataforma de servicios, que incluye la idea de la interacción entre las personas y los sistemas en el desarrollo de una actuación administrativa. El análisis de esta cuestión constituye el objeto de esta comunicación.

2. CONTEXTUALIZACIÓN Y DELIMITACIÓN CONCEPTUAL

La reserva de humanidad, entendida con la proscripción de la automatización en determinadas actuaciones o fases de los procedimientos administrativos, forma parte de un concepto más amplio relativo al gobierno de los sistemas decisorios híbridos.

¹ Este trabajo se inscribe en el proyecto de investigación La actuación administrativa automatizada: condiciones, límites y garantías jurídicas (PID2020-116855RB-100) financiado por la Agencia Estatal de Investigación (AEI/10.13039/501100011033).

Cuando el tema se circunscribe al ámbito de las actuaciones administrativas regidas por el art. 41 de la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público (LRJSP), el concepto de sistema decisorio híbrido abarca toda actuación o fase de la misma, en la que participan, de forma conjunta, sistemas informáticos y personas físicas. En ese contexto, el gobierno del sistema híbrido establece qué actuaciones o trámites han de ser ejecutadas por personas y cuáles por sistemas, atendiendo las exigencias del ordenamiento jurídico.

Dentro del conjunto de alternativas, el gobierno puede implicar:

- Que la decisión se reserve por completo a las personas (PONCE, 2019)
- Conservar la presencia humana en una parte del proceso decisorio

Human-in-the-loop algorithms (el sistema ofrece un elemento de juicio),

Human-on-the-loop algorithms (el sistema realiza una recomendación)

- Excluir a la persona de la decisión: human-out-the-loop algorithms².

Contrario a lo que podría pensarse, esta cuestión no es novedosa, ni está asociada a la utilización de las técnicas de inteligencia artificial; por el contrario, es un asunto que se discute desde mediados del siglo XX (COGLIANESE y HEFTER, 2022: 899 y ss), especialmente en el ámbito de la seguridad. Pese a que son cuestiones antiguas, no se han identificado evidencias acerca de que se consideren plenamente zanjados algunos puntos cruciales del gobierno, tales como, los criterios que rigen las condiciones de interacción o colaboración hombre-sistema. Por lo pronto puede adelantarse que dichas condiciones se fijan en dos planos: normativo u organizativo. En esta comunicación se realizará una breve referencia al primero, para centrar la atención en el segundo.

3. CONDICIONES DE COLABORACIÓN EN EL PLANO NORMATIVO

En el ordenamiento jurídico español pueden identificarse distintas modalidades de colaboración entre los sistemas y las personas, que parten de una premisa: a día de hoy, el concepto de órgano garantiza la presencia de la persona física en las decisiones administrativas (LRJSP; art. 5.1), toda vez que esta ficción, clave para la imputación de la actuación a la Administración, alude a la existencia de un conjunto de medios personales y materiales, articulados en torno a la persona del titular, que ejercitan competencias (SANTAMARIA PASTOR, 1984). Así mismo, la propia redacción del art. 41 LRJSP refuerza la presencia humana en el proceso decisorio, al prever la participación de varios órganos (ergo, varias personas titulares de los mismos) en la actuación automatizada, de donde se sigue que, aún en los casos de plena automatización, no puede hablarse de una ausencia absoluta de personas en la actuación administrativa formalizada.

Con esta idea en mente y, a título meramente ejemplificativo, cabe destacar que el art. 130 del Real Decreto Legislativo 8/2015, de 30 de octubre, por el que se aprueba el texto

² Las nociones de human in the loop (HITL), “on the loop” (HOTL) y “out the loop” (HOOTL) provienen de la ciencia de los ordenadores y no son unívocas. Normativamente tampoco se les ha atribuido un único significado, tal y como puede apreciarse en el análisis de la COM/2019/168 final y del Libro Blanco sobre la inteligencia artificial (COM/2020/65 final/2). Así las cosas, se propone esta diferenciación que distingue tres situaciones con efectos jurídicos distintos: que el resultado del sistema se tenga como un elemento de juicio dentro del proceso decisorio, constituya, en sí mismo, una recomendación o sea el acto resolutorio.

refundido de la Ley General de la Seguridad Social (LGSS) consagra la adopción y notificación de resoluciones de forma automatizada para la protección del desempleo, prestaciones del sistema, afiliación y cotización (out the loop), excepto para las no contributivas, que se someten a reserva de humanidad.

De otra parte, en la normativa procedimental general, se prevé la utilización exclusiva de sistemas (out the loop) esencialmente en actuaciones materiales [remisión de anuncios al BOE (LPAC³; Disposición adicional tercera)], actos de conocimiento con efectos jurídicos limitados o concretos [constancia de presentación en el registro electrónico (LPAC; art. 16.3), emisión de copias auténticas (LPAC, art. 27.1 y RAFESP⁴; art. 48.2), verificación de condiciones de representación (RAFESP; art. 35)]⁵ o interadministrativos [cesión de datos (RAFESP; art. 61.5)]. En la normativa procedimental sectorial, está prevista la adopción de decisiones, sin intervención humana, para iniciar actuaciones⁶, en la fase de instrucción⁷ (on the loop) y en la de resolutive⁸.

Cualquier sistema de detección de comportamientos sospechosos que puedan indicar la existencia de actividades que puedan tipificarse como infracciones administrativas, puede constituir prueba indiciaria que sustente la apertura de un procedimiento de inspección (in the loop).

Se infiere de las disposiciones precitadas que el legislador ha admitido el uso de sistemas en distintas fases del procedimiento: inicio (actas de inspección automatizadas), instrucción (valoración de ofertas), terminación (declaración de devolución tributaria). Así mismo, permite prescindir de la intervención humana en actos de conocimiento y con efectos interadministrativos. Finalmente remarca la necesidad de contar con el razonamiento humano en procedimientos para la concesión de prestaciones a personas vulnerables. A partir de estos ejemplos puede concluirse que las condiciones de interacción se fijan en función a los bienes jurídicos afectados y el alcance de los efectos jurídicos de los actos administrativos: si se está en presencia de actos de conocimiento o interadministrativos, así como en actuaciones que concluyan con efectos favorables, se permite la prescindencia de la intervención humana.

Por el contrario, atendiendo lo previsto en el apartado 26 (principios de proporcionalidad e inocuidad) de la Recomendación sobre la ética de la inteligencia artificial de la UNESCO

3 Ley 39/2015, de 1 de octubre, del Procedimiento Administrativo Común de las Administraciones Públicas.

4 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

5 También está prevista, con estos efectos, la emisión de certificados tributarios (art 73.2 del Real Decreto 1065/2007, de 27 de julio, por el que se aprueba el Reglamento General de las actuaciones y los procedimientos de gestión e inspección tributaria y de desarrollo de las normas comunes de los procedimientos de aplicación de los tributos).

6 El art. 53.1.a) del Real Decreto Legislativo 5/2000, de 4 de agosto, por el que se aprueba el texto refundido de la Ley sobre Infracciones y Sanciones en el Orden Social.

7 El art. 159.6.d) de la Ley 9/2017, de 8 de noviembre, de Contratos del Sector Público (LCSP) prescribe para la valoración automatizada de ofertas en el procedimiento abierto simplificado.

8 El artículo 100.2 de la Ley 58/2003, de 17 de diciembre, General Tributaria (LGT) y el artículo 124 de su Reglamento General de Procedimientos prevén la resolución automatizada en procedimientos que conduzcan a devolución.

(2021), la reserva de humanidad se impone en las actuaciones que impliquen decisiones de vida o muerte o conduzcan a la producción de efectos irreversibles o de difícil reversión. En el mismo sentido, el Proyecto de Reglamento de IA en tramitación⁹ define las prohibiciones en términos de usos con efectos desfavorables (art. 5.1.c) o que impliquen el tratamiento de datos personales sensibles (arts. 5.1.d, 5.2 y 5.3).

La reserva de humanidad entonces, también se construye desde los criterios de bienes jurídicos y efectos: en cuanto a éstos, si las actuaciones conducen a la producción de efectos de difícil o imposible reparación o tienen un carácter desfavorable y desproporcionado, ha de reservarse la decisión final a la persona; lo mismo que si la actuación versa sobre bienes jurídicos, tales, como la protección de datos personales especialmente sensibles.

Entre estas dos alternativas extremas (excluir a los sistemas o a las personas), se sitúan un conjunto de posibilidades de interacción entre los sistemas y la intervención humana. Este concepto, que aparece recogido en el art. 22 del Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (RGPD), también expresa los criterios de referencia de bien (protección de datos personales) y efecto jurídicos (generen efectos jurídicos o sustancialmente equivalentes). El alcance de la intervención, como garantía, se precisa en el WP251¹⁰, aludiendo a una finalidad revisora y de control¹¹, sobre la que también se afirma el carácter de mínimo¹².

En suma, el legislador ha empleado los criterios de bien y efectos jurídicos para definir las condiciones de colaboración entre los sistemas y las personas, sin precisar, de manera expresa, el alcance de la intervención humana para las denominadas, a efectos de esta comunicación, modalidades intermedias. Es la labor interpretativa del Grupo de trabajo del art. 29 la que permite vincular la intervención con el ejercicio de una función revisora, que se extiende sobre todos los actos, sean o no recurribles, si bien esta garantía se circunscribe a los tratamientos de datos personales y a las solicitudes de intervención a petición del titular de los datos.

9 COM (2021) 206: 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.

10 GRUPO DE TRABAJO DEL ART. 29. Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679. Revisadas y aprobadas el 6 de febrero de 2018,

11 Toda revisión debe ser llevada a cabo por una persona con la autorización y capacidad adecuadas para modificar la decisión. El revisor debe llevar a cabo una evaluación completa de todos los datos pertinentes, incluida cualquier información adicional facilitada por el interesado. A su turno, los capítulos II.6.e) y III. C de la Declaración Europea sobre los Derechos y Principios Digitales para la Década Digital (2023/C 23/01) fijan como objetivos de la intervención la salvaguardia de la seguridad, los derechos humanos y la protección del trabajo justo y equitativo.

12 Este carácter de mínimo resulta discutible, toda vez que la propia redacción del art. 22.3 RGPD descarta la existencia de un régimen unitario de garantías, independiente de la fuente de legitimación, al vincular el conjunto de mínimos a fuentes de legitimación distintas a la habilitación normativa. En sentido contrario PALMA ORTIGOSA (2019:20).

Estas dos limitaciones de la garantía (subjetivas, objetivas y formales) demandan una reflexión más general de la institución, que permita delimitarla como garantía procedimental de cualquier interesado en una actuación administrativa.

4. CONDICIONES DE COLABORACIÓN EN EL PLANO ORGANIZATIVO

Las modalidades de interacción entre las personas y los sistemas en el desarrollo de actuaciones administrativas constituyen una expresión del ejercicio de la potestad organizativa de los órganos, teniendo en cuenta que la automatización es sólo una forma o modalidad de tramitación electrónica, que no incide en la distribución de las competencias, ni puede afectar el conjunto de derechos de los interesados en el procedimiento (LPAC; art. 53).

El ejercicio de esta potestad ha de regirse por las siguientes condiciones y principios:

A. Elección de enfoque y planificación

Una decisión organizativa sobre la articulación de medios personales y sistemas en el desarrollo de actuaciones administrativas puede adoptarse desde dos enfoques: el primero, que se denominará a los efectos de este trabajo, pro-técnica (HUQ, 2020) (CERRILLO, 2020), propugna por definir el gobierno desde la limitación de los sistemas, de tal suerte que la intervención humana sólo se justifica si el problema no es resoluble técnicamente. Un enfoque alternativo sería el humanista, que promueve la definición del gobierno desde la limitación del razonamiento humano: sólo deben emplearse sistemas que introduzcan mejoras verificables sobre el desarrollo de las actuaciones (COGLIANESE y HEFTER, 2022: 915) y siempre que su impacto residual sea asumible jurídicamente.

Partiendo de un enfoque humanista, las decisiones algorítmicas pueden introducir mejoras, respecto del razonamiento humano, en términos de: capacidad, eficiencia, consistencia y velocidad. Todos estos atributos sustentan una solución de tipo “human out the loop” en actuaciones masivas, en las que la oportunidad resulte un factor determinante de la eficacia, se ejerciten potestades regladas y los objetivos de consistencia y no discriminación sean los más relevantes (CROOTOF y KAMINSKI:2023, 463) (HUQ, 2020:636) (COGLIANESE y LAI, 2022, 1306) (BINSS, 2022: 204).

De otra parte, las mejoras en el ámbito del ejercicio de las potestades discrecionales, resultan menos generalizables. Su entidad concreta dependerá, bien de la consagración normativa¹³ o del grado y modalidades de la discrecionalidad: en el campo de la discrecionalidad, en sentido pleno, la reserva de humanidad opera con absoluta eficacia, mientras que, en el ejercicio de potestades discrecionales, más o menos regladas, o de discrecionalidad técnica, la reserva de humanidad, si la hubiere, está modulada.

En todo caso, teniendo en cuenta que resulta altamente improbable que se puedan definir, *a priori*, las reglas de resolución de todas y cada una de las casuísticas que se someten a consideración de la Administración en el ejercicio de tales potestades (BINSS, 2022:202) y qué ésta no puede abstenerse de ejercitar sus competencias so pretexto de que los hechos o argumentos no encajan en las reglas programadas en los sistemas, surge naturalmente el

¹³ La norma alemana de procedimiento (§ 35a VwVfG.) prohíbe la automatización en el ejercicio de potestades discrecionales.

derecho a obtener la intervención humana cuando los argumentos (hechos y consideraciones jurídicas) que se pretenden hacer valer ante la Administración pueden calificarse, como atípicos y, en esa medida, no programables.

Definido el marco material general dentro del cual puede verificarse que la introducción de sistemas puede mejorar el desempeño del razonamiento humano, ha de procederse a la fijación de los objetivos que persiguen las mejoras: servicio a los ciudadanos y funcionarios (LJRSP, art. 3.a), simplicidad y claridad de los trámites (LRSSP; art. 3.n), facilidad de uso (RAFESP; art. 2.c), racionalización y agilización (LRJSP; art. 3.d) y eficacia en el cumplimiento de los objetivos (LRJSP; art. 3.h).

La interacción colaborativa entre personas y sistemas en la Administración no puede resolverse de forma aislada, sino que ha de definirse en el marco de un proceso de planificación global (COGLIANESE y LAI, 2022: 1334) (CROOTOF y KAMINSKI, 2023:488-491) con participación de la ciudadanía (GREEN, 2022: 11) (COGLIANESE y HEFTER, 2022: 920). Una técnica al servicio de este objetivo de planificación, es la evaluación de impacto. En el caso del uso de los sistemas, dicha evaluación debe considerar, como mínimo, los siguientes aspectos:

Crterios	Subcriterios	Impacto
Naturaleza	Sujetos afectados Datos Efectos Bienes jurídicos comprometidos Potestad ejercitada Técnica empleada	Menores, vulnerables, riesgo de exclusión Sensibles, personales, no personales Declarativos, constitutivos, favorables, desfavorables DFFF, derechos sociales, derechos legales Discrecionales, regladas IA simbólica, machine learning, etc.
Contexto	Voluntariedad u obligatoriedad	
Alcance	Material (ad extra o ad intra) Temporal (revisable periódicamente, con vocación de permanencia)	
Finalidad del uso concreto Finalidad de la actuación	Elemento de juicio, Recomendación Decisión Consistencia, no discriminación, justicia individual	

B. Delimitación de la función correctiva

Siguiendo lo indicado en el WP 251, a propósito de la interpretación del art. 22 RGPD, el ejercicio de la función correctiva implica que el sujeto está autorizado y es competente para modificar la decisión automatizada, ésto es, comprende el funcionamiento del sistema al punto que puede advertir la existencia de un error y corregirlo¹⁴.

El error en el resultado que arroje el sistema puede obedecer a varios factores: (a) datos (los incluidos no reflejan la realidad, bien porque son erróneos, están desactualizados, son

14 AZIZ (2020: 670) considera que la función correctiva también se ejercita mejor por otros sistemas.

insuficientes o no son relevantes); (b) regla (la propia regla que sigue el sistema no se ajusta al ordenamiento, desde el origen, es decir, el diseño es erróneo, la manera cómo opera el algoritmo ha modificado el diseño original y su deriva no se ajusta a la legalidad o la regla es correcta para la generalidad de los casos, pero no para situaciones singulares); (c) el resultado se ajusta a la regla pero no es razonable jurídicamente o, concretamente, tiene un sesgo (PEETERS, 2020:508).

En cuanto la función de control, la doctrina destaca varias limitaciones físicas y organizativas para su desarrollo (CROOTOF y KAMINSKI:2023, 469) (COGLIANESE y HEFTER, 2022: 919) (GREEN, 2022:11), dentro de las cuales destaca el fenómeno contrastado de la complacencia (tendencia humana a seguir la recomendación del sistema) (PEETERS, 2020: 516¹⁵) (CROOTOF y KAMINSKI, 2023: 469) (GREEN, 2022:8).

Definir un modelo de gobierno de un sistema decisorio híbrido, desde las personas y no desde la tecnología, implica reconocer el efecto de estos comentarios críticos sobre el ejercicio de la función correctiva, para tratar de solventarlos, partiendo de la premisa de que la función correctiva: opera como un continuo (se ejercita en varios momentos y los resultados pueden desencadenar varias acciones correctivas) (PEETERS, 2020:512), requiere definir claramente el objeto de control para, a partir del propósito (CROOTOF y KAMINSKI,2023: 487), concretar el tipo de intervención humana que está en condiciones de alcanzarlo de acuerdo con su formación (PEETERS, 2020: 517), su conocimiento del sistema - comprensión del funcionamiento y uso- (GREEN, 2022:6) y su posición dentro de la organización (CROOTOF y KAMINSKI,2023: 502), tomando en consideración que el alcance difiere según la modalidad de la intervención:

	Diseño	Inicio	Instrucción	Resolución	Recurso
Sesgos	Cantidad y calidad de los datos Objetivos claros de la decisión		Elemento de juicio (in the loop) hay que hacer una valoración con arreglo a la sana crítica.	Recomendación (on the loop) Es un informe que puede ser preceptivo, pero no vinculante. Propuesta de resolución sujeta a traslado.	Quien conoce del recurso debe estar en capacidad de detectar el sesgo
Exactitud	Calidad de los datos	Previsualización de la instancia	Instructor detecta la inexactitud en el procedimiento.	Se aparta por inexactitud Se controla mediante el traslado de la propuesta.	Recurso
Error en la traducción de la regla	Objetivo crucial del diseño: traducir en		Puede ser detectado, tanto si el out put es un	Detectado por el activo.	Detectado por el competente para conocer del recurso.

15 Los algoritmos adquieren cada vez más poder porque los operadores los ven como racionales, científicos y neutrales.

	reglas las exigencias normativas		elemento de juicio, como si se trata de una recomendación.		
Error por ser atípico	Un buen diseño puede limitar la posibilidad de que surja el atípico	El hecho es declarado por el propio interesado. Art. 150.7 ley federal tributaria reconoce el derecho del contribuyente a solicitar la intervención	Instructor detecta error a la vista de información adicional aportada por el interesado.	El interesado aporta información durante el traslado de la propuesta de resolución y el activo detecta el carácter atípico	Detectado por el competente para conocer del recurso.

El control sobre el funcionamiento del sistema no es una función puntual, sino que debe llevarse a cabo a todo lo largo de la vida útil del sistema (desde su diseño hasta la cancelación de los datos de los que se nutre). Se trata de un proceso colaborativo, iterativo, que involucra a los interesados y demanda una acción coordinada entre órganos¹⁶ integrados por personal formado GREEN, 2022: 11) (COGLIANESE y HEFTER, 2022: 920). Así mismo, parte del control consiste en implementar mecanismos de rendición de cuentas ante el público y organizaciones democráticas (GREEN, 2022:11)

El control debe ejercitarse, aún si es el sistema es desarrollado por un tercero: contractualmente ha de preverse la publicidad de los algoritmos, datos y condiciones de entrenamiento, así como la adhesión a guías éticas de desarrollo (COGLIANESE y LAI, 2022: 1335)

5. REFERENCIAS

- Binns, R. (2022). Human Judgment in algorithmic loops: Individual justice and automated decision-making. *Regulation & Governance*, 16(1), 197-211.
- Cerrillo Martínez, A. (2020). ¿Son fiables las decisiones de las Administraciones públicas adoptadas por algoritmos?. *European review of digital administration & law*, 1(1-2), 18-36.
- Coglianese, C., & Hefter, K. (2021). From Negative to Positive Algorithm Rights. *Wm. & Mary Bill Rts. J.*, 30, 883.
- Coglianese, C., & Lai, A. (2021). Algorithm vs. algorithm. *Duke LJ*, 71, 1281.
- Crotoof, R., Kaminski, M. E., Price, W., & Nicholson, I. I. (2023). Humans in the Loop. *Vand. L. Rev.*, 76, 429.
- Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review*, 45, 105681.
- Huq, A. Z. (2020). A right to a human decision. *Va. L. Rev.*, 106, 611.
- Palma Ortigosa, A. (2019). Decisiones automatizadas en el RGPD. El uso de algoritmos en el contexto de la protección de datos. *Revista general de Derecho administrativo*, 50
- Peeters, R. (2020). The agency of algorithms: Understanding human-algorithm interaction in administrative decision-making. *Information Polity*, 25(4), 507-522.

16 Una vez detectado el error, se da a conocer al auditor y éste pone en conocimiento del hecho al órgano responsable de la especificación, programación, calidad del sistema, para que se introduzcan los cambios pertinentes.

Ponce Solé, J. (2019). Inteligencia artificial, Derecho administrativo y reserva de humanidad: algoritmos y procedimiento administrativo debido tecnológico. *Revista general de Derecho administrativo*, 50, 1-52.

Santamaría Pastor, J (1984). *Revista española de Derecho Administrativo*, 40.