



22 de octubre de 2026

Alfredo Gil Laso *

Recibir BOLETÍN ELECTRÓNICO

1

Algocracy: The Threats of Automated Decision-Making

Abstract:

This article analyses the main legal, strategic, ethical, and social risks arising from the growing trend of delegating decisions (or the analyses that lead to decisions) to algorithms and artificial intelligence, especially in government, public administration, and security and defence contexts. The hypothesis of a form of government known as algocracy, in which algorithmic logic partially or totally replaces human judgment, serves as an analytical tool for the current trend of implementing these technologies in decision-making mechanisms. The article examines the challenges this transformation entails and proposes mechanisms for responsible and ethical use. It emphasizes that effective human control must be the guiding principle in any automated system, ensuring traceability, transparency, and the protection of fundamental rights.

Keywords:

Automation, Artificial Intelligence, Civic Apathy, Algorithmic Opacity, Transparency, Digital Literacy, Algocracy.

Cómo citar este documento:

GIL LASO, Alfredo. *Algocracia: los peligros de las decisiones automatizadas*. Documento de Opinión IEEE 106/2026.

Introducción

La sociedad moderna se ha caracterizado por un progreso sin precedentes en su desarrollo tecnológico. Desde la Guerra Fría y su competición entre bloques por demostrar quién estaba más avanzado, la sociedad se ha desarrollado mediante procesos tecnológicos de complejidad iterativa que han dado forma a la realidad contemporánea. En este contexto, era cuestión de tiempo que la capacidad de la mente humana para procesar datos y resolver problemas fuera superada por la necesidad de disponer de resultados rápidos, fiables y precisos a una escala inabarcable. Así surgieron los primeros sistemas de razonamiento mediante algoritmos que, progresivamente, evolucionaron hacia las modernas inteligencias artificiales (IA) y las redes neuronales. Con estos sistemas se ha alcanzado tal grado de refinamiento analítico que logran imitar, con un mimetismo asombroso, el comportamiento cognitivo del ser humano.

De esta forma, los algoritmos, las IA, las redes neuronales y las tecnologías afines se han ido integrando en las últimas décadas en todos los aspectos de la vida moderna. Más allá de los asistentes virtuales o los simuladores de conversación —*chatbots*— a los que exponemos nuestra curiosidad mundana, hemos integrado estas tecnologías de forma discreta en buscadores de Internet, plataformas de contenido, predicciones de consumo, navegación GPS, automatización industrial, *smartphones*... Están tan integrados en nuestras vidas que, realmente, participan en casi todos los aspectos importantes de nuestra sociedad.

Incluso existe la tendencia a realizar consultas de alta complejidad técnica por parte de profesionales cualificados para facilitarse el trabajo. Sirvan como ejemplos: las búsquedas de jurisprudencia de jueces y abogados; el uso por parte de políticos, publicistas y estadistas para analizar patrones sociales e intereses de la población; o los programas de reconocimiento facial y de huellas dactilares que agilizan la labor policial.

Desde una perspectiva de eficiencia, es legítimo emplear estos mecanismos para acelerar la toma de decisiones, pero esto abre el camino a la verdadera cuestión que se plantea en este artículo: ¿dónde está el límite que debemos poner a la confianza depositada en estos sistemas?

Existe una tendencia cada vez más clara de apoyarnos en exceso en los algoritmos, obviando el hecho de que la responsabilidad última es del ser humano y no de la IA que procesa la información. Si depositásemos nuestra confianza en los algoritmos hasta el punto de no cuestionar su criterio, podríamos recaer en un proceso de «apatía cívica» que desembocaría en uno de los pilares centrales de este artículo: un gobierno regido directa o indirectamente por las decisiones de una IA. A esa forma de gobierno es a lo que se llama «algocracia».

Nunca ha existido algo semejante, ni parece viable a corto plazo, pero nos sirve como mecanismo especulativo para analizar hasta dónde puede llevarnos nuestra cada vez mayor dependencia de los algoritmos y las IA. No en vano, poner en manos de máquinas o *softwares* ciertos aspectos de la vida y del gobierno puede acarrear importantes peligros estratégicos, éticos y sociales que ya están siendo ampliamente documentados. Por tanto, este artículo pretende analizar qué riesgos asumimos cuando delegamos parte del poder decisorio en los algoritmos.

La «algocracia»: una tendencia no tan nueva

En este punto es cuando entramos de lleno en el terreno de lo meramente especulativo, como corresponde a la hipótesis del concepto de «algocracia». Básicamente, se trataría de considerar qué pasaría si, con el tiempo, nuestra confianza en las IA pasase de lo meramente consultivo a la delegación de tareas sensibles, complejas o trascendentales.

En estas circunstancias, con el debido desarrollo tecnológico, podríamos llegar a que una inteligencia artificial lo bastante sofisticada acaparase cuotas de decisión que, *de facto*, supusiesen el control de todos los aspectos que rigen la sociedad. Podríamos decir que esa IA es quien realmente gobernaría. Sería un riesgo significativo si se combina con el desinterés de la población por su propia gobernanza («apatía cívica») —que es un fenómeno ya existente en la sociedad por otros motivos—. Se podría considerar, en ese caso, que el sistema de gobierno habría pasado a ser una nueva forma de totalitarismo por desistimiento social: una dictadura «algocrática»¹.

¹ VILLARROEL SOTO, Raúl. «Algocracia. Decisiones políticas y representación ciudadana en una gobernanza algorítmica», *Revista de Filosofía "Otro siglo"*. 27/10/2023, pp. 193-212. ISSN 0719-921X. DOI: 10.5281/zenodo.10451886.

Técnicamente se podría definir la «algocracia» como un fenómeno político, aún hipotético, por el cual los algoritmos, especialmente aquellos basados en inteligencia artificial y aprendizaje automático, se convierten en decisores activos sobre la gobernanza de un territorio, región o Estado. Para ello, desplazarían o condicionarían la soberanía nacional, su autonomía estratégica y la capacidad de decisión democrática en favor de sistemas automatizados gestionados por estructuras tecnocráticas que, probablemente, serían de gestión privada o transnacionales.

Desde una perspectiva filosófica, se trataría de una mutación del paradigma de autoridad, una iteración evolutiva conceptual que implicaría la reconfiguración de las nociones de poder, responsabilidad y subjetividad. Se plantearía el desafío de cómo preservar la dignidad del ser humano cuando es una máquina lógica la encargada de tomar decisiones que pueden resultar trascendentales para el individuo. Sería el paradigma último de la racionalidad instrumental de Max Weber, corriendo el riesgo de conducir a la sociedad a una «jaula de hierro» de eficiencia deshumanizada².

A primera vista puede parecer solo una nueva forma de tecnocracia o de democracia digital, o incluso una nueva fórmula de burocracia virtual, pero la «algocracia» se distingue de estos conceptos por su naturaleza absolutamente automática, su lógica algorítmica carente de sesgos humanos (aparentemente) y su impacto en la soberanía.

Se podrán definir varios tipos de «algocracia»: basándonos en sus fines, quién controle su *software*, si existe connivencia social para su instauración o si ha sido un proceso progresivo e imperceptible. Sin embargo, algunos autores consideran que la clasificación más relevante sería en función de las consecuencias implícitas de los procesos de decisión, esto es, si sus indicaciones se vuelven directamente vinculantes o si es necesaria la ratificación por parte de un humano que se responsabilice de la misma³.

No obstante, el ejercicio mental propuesto en este artículo no se trata por completo de una especulación. En la actualidad ya se están implementando mecanismos decisorios en los cuales se emplean algoritmos, IA y redes neuronales para la toma de decisiones relevantes desde un punto de vista gubernamental. Algunos ejemplos son la

² GIL VILLEGAS, Francisco. «El concepto de racionalidad en la obra de Max Weber», *Revista Mexicana de Ciencias Políticas y Sociales*, 31(119). 26/11/2019. DOI: 10.22201/fcpys.2448492xe.1985.119.72143.

³ CELESTE DANESI, Cecilia. «El peligro de automatizar decisiones con IA sin control democrático», *OnIA, IA Ética*. Buenos Aires, 15/6/2025. Disponible en: www.elnacional.cat/oneconomia/es/on-ia/peligro-automatizar-decisiones-con-ia-sin-control-democratico_1434472_102.html (consultado el 15/9/2025).

automatización de la revisión de documentos judiciales en la UE, la predicción de fallos en puentes y carreteras en EE. UU., la atención automatizada mediante *chatbots* en muchos municipios y organismos públicos del mundo, la optimización de concesiones crediticias, la recepción automática de denuncias contra la policía mexicana o la evaluación de recursos humanos en algunas agencias de Reino Unido⁴.

La automatización de lo público: ventajas o espejismos

Cuando hablamos de qué ventajas aporta a la administración y a la gobernanza el uso de las IA no solo debemos tratar el tema desde una perspectiva de eficiencia operativa. El uso de herramientas de análisis de datos aporta una mejora sustancial en la toma de decisiones, puesto que facilita la identificación de patrones y tendencias, permitiendo orientar las políticas públicas en base a evidencias lógicas.

Asimismo, una IA no tiene intereses económicos propios —aunque más adelante se matizará esta afirmación—, por lo que su uso permite una mayor transparencia y trazabilidad en la toma de decisiones. Esto minimiza la opacidad de la gestión pública y facilita la rendición de cuentas ante decisiones erróneas o subrepticias. En este sentido, el uso de las IA permitiría disminuir el riesgo de fraude gracias a mecanismos de detección temprana de irregularidades, monitorizando automáticamente transacciones y expedientes de gasto en una búsqueda constante de anomalías que evidencien posibles infracciones⁵.

Además, el uso de las IA para atender las peticiones y necesidades de los ciudadanos permitiría una atención más profesionalizada, ágil e inmediata, mejorando la experiencia del administrado y permitiendo incluso la segmentación inteligente de beneficios y recursos. Esto incrementaría la confianza en la Administración puesto que, por un lado, se percibe como un síntoma de modernización y profesionalización, pero, por el otro, reduciría los errores humanos y el riesgo de arbitrariedad administrativa⁶. En definitiva, se percibiría como una Administración más profesional, equitativa, automática y eficiente,

⁴ WWW.GOBIERTO.ES. «Inteligencia Artificial en la administración pública: casos de uso». Disponible en: https://www.gobierno.es/blog/20190920-inteligencia-artificial-administracion-publica?gad_source=1&qad_campaignid=22714986660&qclid=Cj0KCQjw8p7GBhCjARIsAEhghZ1p5-7ST3sXHWlU7FDRwnwHDhWJkm583icbOYkuLEL6iw90QovEWlaAgcPEALw_wcB (consultado el 15/9/2025).

⁵ CAIZA, Gustavo *et al.* «Navigating Governmental Choices: A Comprehensive Review of Artificial Intelligence's Impact on Decision-Making», *Revista Informatics*, 11, 64, MDPI, 4/9/2024. DOI 10.3390/informatics11030064.

⁶ BABŠEK, Matej *et al.* «Artificial Intelligence Adoption in Public Administration: An Overview of Top-Cited Articles and Practical Applications», *Revista AI*, 6, 44, MDPI, 21/2/2025. DOI: 10.3390/ai6030044.

con la capacidad de reaccionar a las necesidades de la sociedad que administra con la inmediatez de la que muchas veces carece la burocracia actual.

Pero no todo son ventajas en el uso de las IA para la administración y el gobierno. El principal problema al que se enfrenta una sociedad al implementar estos sistemas es el error de considerar que los algoritmos, por estar basados en lógica matemática y estadística, están exentos de sesgos humanos y son completamente objetivos e imparciales en sus afirmaciones. Esta idea se llama «espejismo de la neutralidad algorítmica»⁷ y ha sido ampliamente analizada por el mundo académico, que ha comprobado empíricamente que los algoritmos pueden llegar a reproducir, a amplificar o incluso a crear nuevas formas de discriminación, dependiendo únicamente de cómo se diseñan, entrenan y aplican. En este punto es necesario aclarar que las IA son desarrolladas por entidades privadas con intereses económicos y políticas institucionales propias, que no tienen por qué estar alineados con los del gobierno o la sociedad que utiliza esos sistemas.

También se debe considerar la posibilidad de operar en un entorno ya sesgado por desigualdades sociales, diferencias étnicas o polarización política. Un ejemplo notable es el de los sistemas de reconocimiento facial, en los cuales se ha demostrado que aquellos modelos eminentemente entrenados con personas blancas sufren una mayor tasa de fallos a la hora de identificar a personas negras —es un hecho extrapolable a cualesquier otra etnia—. Este es un error que evidencia un sesgo debido a la limitación o segmentación de los datos disponibles, pero ilustra conceptualmente cómo un análisis correctamente ejecutado puede cometer errores debido a los datos de partida⁸.

Pero no debemos considerar que el «espejismo» se produce únicamente por fallos humanos —involuntarios o negligentes— al alimentar el sistema. No olvidemos que Google fue multada en 2017 y de nuevo en 2025 por favorecer sus propios productos en los motores de búsqueda, demostrando que se pueden introducir sesgos en los algoritmos en favor de su creador y/o propietario. Como se indicó en la introducción de

⁷ PHILLIPS-BROWN, Milo. *Algorithmic neutrality*. University of Oxford, v4, 22/7/2025.
DOI: 10.48550/arXiv.2303.05103.

⁸ STINSON, Catherine. *Algorithms are not neutral: Bias in collaborative filtering*. *AI Ethics* 2, 763–770, 31/1/2022.
DOI: 10.1007/s43681-022-00136-w.

este apartado, es cierto que los algoritmos no tienen intereses propios, pero sí pueden tener un sesgo externo introducido por sus propietarios o diseñadores⁹.

Los peligros de la algocracia

Volvamos a la hipótesis de partida: la «algocracia», un gobierno completamente regido por algoritmos ya sea por la aceptación explícita de los ciudadanos o por una dependencia implícita de los decisores humanos.

En esta hipotética forma de gobierno, lo primero con lo que nos encontraríamos sería un problema de «opacidad algorítmica»: una absoluta falta de comprensión de los intrincados procesos de razonamiento de la inteligencia artificial gobernante por parte de las personas. Este problema podría derivar en un fenómeno de ineptitud política, dotando a los responsables últimos de las decisiones —si es que siguiera habiéndolos— de argumentos para eludir su responsabilidad cuando los resultados fuesen negativos¹⁰.

Tal y cómo hemos visto, es posible inducir sesgos concretos en las decisiones de estos sistemas simplemente aportando datos estadísticos incompletos o preseleccionados, ya sea por indolencia, incompetencia o corrupción. Esto podría llevar a tomas de decisiones en contra de lo que se considera justo, ético o lícito, generando efectos indeseados como podría ser la discriminación automatizada de colectivos concretos¹¹.

Además, es necesario prestar atención a la dependencia de los algoritmos y las IA en la toma de decisiones críticas para la supervivencia del Estado. En consecuencia, estos sistemas se convierten en objetivos prioritarios de eventuales enemigos. Si los responsables humanos pierden la capacidad de tomar decisiones de forma autónoma, la dependencia tecnológica se habría convertido en una vulnerabilidad estratégica demasiado crítica¹².

⁹ XAVIER, Bibin. «Biases within AI: challenging the illusion of neutrality», *AI & Society*, Volume 40, 13/6/2024, pp. 1545-1546. DOI: 10.1007/s00146-024-01985-1.

¹⁰ SZLAJEN, Fishel. «Opacidad algorítmica e implicancias bioéticas», *InfoBAE*. 16/11/2024. Disponible en: <https://www.infobae.com/opinion/2024/11/16/opacidad-algoritmica-e-implicancias-bioeticas/> (consultado el 15/9/2025).

¹¹ ARAYA PAZ, Carlos. «Transparencia algorítmica: ¿un problema normativo o tecnológico?», *CUHSO* (Temuco). ISSN-e 2452-610X, vol. 31, n.º 2. Diciembre 2021, pp. 306-334. DOI: 10.7770/cuhso-v31n2-art2196.

¹² ROCA RIVERO, Francisco Javier. «La era de la Inteligencia Artificial ya está aquí», *Revista Española de Defensa RED*, n.º 430, año 38. Septiembre de 2025. Disponible en: <https://publicaciones.defensa.gob.es/media/downloadable/files/links/r/e/red-430.pdf>.

Por último, no debemos olvidar la desigualdad digital como brecha social. Este fenómeno se da entre quienes tienen la capacidad de entender, gestionar o controlar los algoritmos y las IA, y los que se limitarían a aceptar las decisiones derivadas de sus análisis, aunque no las comprendan. Con esto se correría el riesgo de que apareciese una nueva casta privilegiada, basada en las competencias tecnológicas, que modificaría el equilibrio social hacia una desigualdad progresivamente más acusada¹³.

Apatía cívica: un nuevo riesgo

La «apatía cívica» que deriva en una «algocracia» no es una idea consolidada sino una noción inferida e incipiente. Se trata simplemente de dos conceptos que se interrelacionan para establecer una idea en su conjunto: una relación causa efecto que, en la situación concreta analizada en este artículo, ofrece un epítome sobre un problema cuya profundidad podría llenar tomos enciclopédicos completos.

Sin embargo, con la información que ya hemos expuesto en este texto, podemos deducir que la «algocracia» por «apatía cívica» sería la consecuencia de la indiferencia, el desinterés o la desmoralización de la ciudadanía hacia los asuntos públicos. En este caso se derivaría de la comodidad de delegar decisiones de gestión y gobierno en sistemas algorítmicos, que se suponen infalibles, pero que en la práctica se percibirían como opacos, incontrolables e inapelables¹⁴.

Las causas podrían ser muy diversas, y se deberían estudiar los casos de forma individual, si se produjeran. Sin embargo, podemos anticipar que los factores educativos y culturales provocarían una falta de comprensión del funcionamiento de los algoritmos que finalmente desembocaría en la aceptación ciega o el escepticismo. En un primer momento, el ciudadano medio pasaría a delegar en los procesos de decisión automática cualquier aspecto de gobierno, deslegitimando la política tradicional y sustituyéndola por una nueva tecnocracia basada en procesos técnicos, impersonales e inhumanos (en el

¹³ SARDEGNA, Paula Costanza y PELIZA, Eleonora. «Desafíos jurídicos ante la gobernanza de datos y la opacidad algorítmica en el empleo», *Revista Internacional y Comparada de Relaciones Laborales y Derecho del Empleo*, volumen 12, n.º 4. Octubre-diciembre de 2024. ISSN 2282-2313.

¹⁴ ALNEMR, Nardine. «Democratic self-government and the algocratic shortcut: the democratic harms in algorithmic governance of society», *Contemporary Political Theory*, vol. 23. 2023, pp. 205-227.
DOI: 10.1057/s41296-023-00656-y.

sentido de artificiales y lógicos). Sería un sistema tan cómodo e incuestionable que generaría el riesgo de desactivación progresiva de la participación ciudadana¹⁵.

Es en este punto donde la desinformación o sobreinformación serían el caldo de cultivo para magnificar los desajustes o polémicas generadas por el sistema, legitimando cualquier corriente de pensamiento contraria al régimen preponderante (aunque sean radicales). Progresivamente, todo ciudadano que percibiera como injusta alguna decisión de la «algocracia» se expondría a sentimientos de desafección hacia el sistema, cayendo en el fatalismo de atribuir todos sus males a la tecnología y desmotivándose respecto a la participación ciudadana en la vida política. Este proceso supondría un peligroso caldo de cultivo para la erosión del contrato social y la progresiva deslegitimación del sistema de gobierno¹⁶.

Implicaciones para la seguridad y defensa

Al margen de la hipótesis planteada, podemos estudiar este campo desde innumerables perspectivas. Sin embargo, ninguna acarrea los riesgos y dilemas morales de delegar decisiones relacionadas con contextos militares y de seguridad nacional¹⁷.

En la actualidad ya se están tomando decisiones estratégicas en base a indicaciones de algoritmos o las IA. A esto se une su creciente presencia como elementos intrínsecos de sistemas de armas, guiado de municiones, procesamiento de inteligencia táctica y sistemas de ciberdefensa¹⁸. Sin embargo, debemos ser conscientes de que cualquier acción en este tipo de escenarios trae consigo una elevada probabilidad de daños a personas, ya sean blancos legítimos o colaterales. Por tanto, cobra importancia la cuestión de la responsabilidad sobre las actuaciones erróneas: si el proceso de decisión se apoya cada vez más en sistemas de elevada complejidad, hasta el punto

¹⁵ GRIMMELIKHUIJSEN, Stephan y MEIJER, Albert. «Legitimacy of Algorithmic Decision-Making: Six Threats and the Need for a Calibrated Institutional Response», *Perspectives on Public Management and Governance*, Volume 5, Issue 3. 28/3/2022, pp. 232–242. DOI: 10.1093/ppmgov/gvac008.

¹⁶ VIEIRA MAGALHÃES, João. «Algorithmic Resistance as political disengagement», *Media International Australia*, vol. 183. 1/5/2022, pp. 77-89. DOI: 10.1177/1329878X221086045.

¹⁷ OLIER, Eduardoy CORCHADI, Juan Manuel. *Inteligencia Artificial: aplicaciones a la Defensa*. Documento de investigación 01/2022. Instituto Español de Estudios Estratégicos. Disponible en: https://www.ieeee.es/Galerias/fichero/docs_investig/2022/DIEEEEINV01_2022_EDUOLI_Inteligencia.pdf (consultado el 10/9/2025).

¹⁸ MEZA, Milton. *Sistemas de armas completamente autónomos: un desafío para la comunidad internacional*. Documento de Opinión 85/2016, IEEE, 18/8/2016. Disponible en: https://www.ieeee.es/Galerias/fichero/docs_opinion/2016/DIEEEE085-2016_SistemasArmas_ONU_MiltonMeza.pdf (consultado el 16/9/2025).

de dificultar la trazabilidad de las decisiones tomadas, ¿cómo se evitarían los vacíos de responsabilidad?^{19,20}

Puede parecer un tema baladí, incluso es tentador aludir a que la responsabilidad última siempre será del decisor humano, pero no es tan sencillo. En un supuesto en el que se tome una decisión errónea, desafortunada, negligente o punible, un auditor tendría que definir no solo quién aceptó un determinado resultado del análisis y actuó en consecuencia, sino también cómo fue el proceso para llegar a esa decisión. La propia naturaleza de los análisis realizados —mediante *softwares*, algoritmos o IA— implica que, para realizar una correcta trazabilidad de los procesos, es necesario recurrir a sistemas análogos, puesto que están diseñados para realizar razonamientos en pocos segundos que resultan imposibles para los seres humanos.

En los casos en que una decisión esté fundamentada en datos de partida o premisas incorrectas, esta trazabilidad se vuelve crítica. Es entonces imperativo determinar si el daño es consecuencia de esos datos o premisas, del propio diseño de los algoritmos o si se trata de un sesgo externo no deseado. Podría incluso ser algo intencionado, puesto que la dependencia tecnológica convierte a estas herramientas en objetivos prioritarios de ciberataques. Al fin y al cabo, el dominio estratégico del ciberespacio se caracteriza por una sucesión de operaciones para limitar, controlar o anular la infraestructura de datos del enemigo, y eso incluye a los algoritmos orientados a la toma de decisiones²¹.

Implementar sistemas que podamos definir como «algocráticos» en seguridad y defensa implicaría una paulatina y significativa pérdida del control humano a favor de una mejora sustancial de la eficiencia, la inmediatez y los resultados. Esto plantea dilemas éticos y estratégicos significativos, erosionando la doctrina clásica de mando y control y sustituyéndola por una serie de sistemas lógicos automatizados que no están limitados por juicios morales.

Como hemos visto, delegar en automatismos actividades tan críticas como la vigilancia, la ciberdefensa o la identificación de objetivos conlleva riesgos específicos. La

¹⁹ NATO Official texts. *Summary of NATO's revised Artificial Intelligence (AI) strategy*. Washington DC, 10/7/2024. Disponible en: https://www.nato.int/cps/en/natohq/official_texts_227237.htm (consultado el 16/9/2025).

²⁰ INSTITUTO ESPAÑOL DE ESTUDIOS ESTRATÉGICOS. «Inteligencia artificial en la geopolítica y los conflictos», *Cuaderno de Estrategia* 226. IEEE, 2024. Disponible en: https://www.ieee.es/Galerias/fichero/cuadernos/CE_226/CE_226.pdf (consultado el 20/9/2025).

²¹ NATO Standardization Office. *Allied Joint Doctrine for Cyberspace Operations*. NATO STANDARD AJP-3.20, Edición A, Versión 1, 29/1/2020.

automatización de decisiones en estos ámbitos es especialmente compleja por el impacto que pueden tener. Cuando la vida, la libertad o los derechos de cualquier persona, incluso un enemigo declarado, estén en juego por la lógica de un algoritmo, es imperativo establecer mecanismos de control humano que resulten eficaces, evitando que el «espejismo de la neutralidad» nos haga considerarlos superfluos o asumibles²².

Gobernanza algorítmica responsable

Es necesario incidir en que el sistema de gobierno basado total o parcialmente en la «algocracia» aún no escapa de lo meramente hipotético, por lo que las propuestas para mitigarlo que vienen a continuación deben interpretarse como teóricas. No obstante, la intención de este artículo es que sirvan de ejemplo para amortiguar los efectos colaterales de delegar en exceso las decisiones en sistemas lógicos automáticos, lo cual sí es una realidad actual.

También es necesario aclarar que los principios rectores con los que se debe gestionar cualquier proceso de toma de decisiones genérico son la transparencia y la responsabilidad. Esto también se aplica al uso de algoritmos o IA con ese fin²³.

Para ello podemos extrapolar los mismos mecanismos de control de decisiones que se deben aplicar a cualquier forma de gobierno: la auditoría y la participación ciudadana. El primero consistiría en procesos sistemáticos orientados a evaluar el funcionamiento de los algoritmos de decisión, verificando su equidad, transparencia, trazabilidad y ausencia de sesgos. Como se indicó en el apartado anterior, esto incluye un problema adicional derivado del volumen de datos y procesos complejos que manejan: solo pueden ser analizados por otro algoritmo o inteligencia artificial. Entraríamos así en un bucle sobre quién audita al algoritmo auditor. El otro método sería el de la participación ciudadana, que podría consistir en paneles o grupos de trabajo de personas no expertas que verifiquen si las decisiones tomadas se ajustan a los criterios éticos y morales de la sociedad administrada. El problema, en este caso, es decidir quién conforma estos grupos de trabajo, si deben ser personas dedicadas profesionalmente a ello o

²² U. S. Department of Defense. *DoD Directive 3000.09: Autonomy in Weapon Systems*. Washington, D. C., 25/11/2023 (primera edición 21/11/2012).

²³ LAMAS LÓPEZ, Francisco y PERALTA GUITERREZ, Alfonso. «Notas sobre el contexto legal internacional y de la UE para el uso militar de la inteligencia artificial». *Boletín de Observación Tecnológica en Defensa BOTD*, n.º 78, 3^{er} trimestre de 2023. Ministerio de Defensa. Disponible en:
https://publicaciones.defensa.gob.es/media/downloadable/files/links/b/o/bot_78.pdf

ciudadanos voluntarios. Ambos modelos ofrecen dificultades a la hora de establecer una auténtica independencia o un interés genuino en el objetivo buscado. Serían mecanismos análogos al de los jurados populares o las mesas electorales²⁴.

Por otro lado, la velocidad a la que se están integrando estos sistemas en todos los aspectos de la vida diaria es abrumadora. Sin embargo, si el proceso de integración fuese lo bastante paulatino como para acompañarlo de políticas de educación digital y alfabetización algorítmica, se podría evitar recaer en la «apatía cívica». Que exista un porcentaje importante de población con una formación mínima que les permita entender los mecanismos de funcionamiento de los algoritmos, las redes neuronales y las IA, permitiría un uso más responsable de estas tecnologías, evitando una brecha social que generaría un nuevo tipo de discriminación. De esta forma se conseguiría que quienes «sufren» las decisiones derivadas de los análisis algorítmicos puedan interpretarlos de forma crítica y constructiva, reduciendo así la asimetría cognitiva entre los administradores y los administrados²⁵.

Por último, no hay que olvidar la importancia de determinar el flujo de responsabilidad. Se debe implementar un marco jurídico y normativo destinado a facilitar la identificación de funciones, la trazabilidad, la «explicabilidad» y los derechos fundamentales en decisiones tomadas mediante sistemas automatizados. Para ello, el concepto de «control humano significativo» —el decisor humano como supervisor con capacidad de veto— permitiría que siempre hubiese un encargado de gestionar las decisiones, evitando la dilución de la responsabilidad cuando los resultados indeseados fuesen punibles^{26,27}.

²⁴ KOENE, Andreas; CLIFTON, Chris; HATADA, Yuko *et al.* *A governance framework for algorithmic accountability and transparency*. European Parliamentary Research Service (STOA), 2019. Disponible en: https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU%282019%29624262%28ANN1%29_EN.pdf (consultado el 17/9/2025).

²⁵ GAGRČIN, Elma; NAAB, Teresa K. y GRUB, Marvin F. *Algorithmic media use and algorithm literacy: An integrative literature review*. New Media & Society, 2024. Disponible en: <https://madoc.bib.uni-mannheim.de/68149/1/gagrcin-et-al-2024-algorithmic-media-use-and-algorithm-literacy-an-integrative-literature-review.pdf> (consultado el 17/9/2025).

²⁶ CHEONG, Boon Chong. «Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making», *Frontiers in Human Dynamics*, vol. 6. 2024. DOI: 10.3389/fhumd.2024.1421273.

²⁷ CAVALCANTE, Paulo. «Artificial Intelligence and Algorithm Accountability: Potential Risks and Proposed Governance Measures», *Blavatnik School of Government Working Paper*. University of Oxford, 2024. Disponible en: https://www.bsg.ox.ac.uk/sites/default/files/2024-09/BSG-WP-2024-60_EN_WorkingPaper_ArtificialIntelligence.pdf (consultado el 17/9/2025).

Una nueva realidad

Durante la redacción de este artículo ha surgido una noticia que hace pensar que el paradigma expuesto en estas líneas está más cerca de hacerse realidad de lo que podría parecer. La noticia indica que el Gobierno de Albania ha implementado una IA llamada Diella («Sol» en albanés) como ministra dedicada a la lucha contra la corrupción, encargándose de todas las licitaciones de la Administración pública albana²⁸.

Aunque se trate de un acto simbólico, este hecho ilustra la cuestión de la creciente capacidad de decisión de los algoritmos, dejando en sus manos tareas tan importantes como la supervisión de los gastos de un país²⁹. No es objetivo de este artículo valorar la calidad democrática de este experimento político, ni los intereses de las empresas que están detrás de su desarrollo (con su propio contrato gubernamental), ni si los miembros del Gobierno se han reservado algún tipo de salvoconducto para introducir sesgos. Lo que sí es necesario analizar en un país que no destaca por su transparencia —puesto 80 de 180 sobre percepción de la corrupción en 2024³⁰— es si existe una jerarquía de responsabilidad ante los eventuales errores que pueda cometer Diella o si, por el contrario, los decisores políticos se limitarán a acogerse al «espejismo de la neutralidad algorítmica», apuntando a su supuesta infalibilidad. Con el tiempo veremos los resultados, la evolución de esta idea y los nuevos problemas que deberán afrontar.

Conclusiones

En este texto se ha planteado la delegación de decisiones en algoritmos como un escenario hipotético a la par que plausible, pero ya en la actualidad se están dando pasos importantes que nos acercan cada día un poco más a esta situación. Esto plantea una serie de riesgos éticos, sociales y estratégicos que sería negligente ignorar, máxime

²⁸ DAVIES, Pascale. «Albania nombra a la primera "ministra" del mundo generada por inteligencia artificial para erradicar la corrupción», *Euronews digital*, sección Negocios Tecnología. 13/9/2025. Disponible en: <https://es.euronews.com/next/2025/09/13/albania-nombra-a-la-primera-ministra-de-inteligencia-artificial-del-mundo-para-erradicar-> (consultado el 18/9/2025).

²⁹ ELECONOMISTA.ES. «Albania nombra a una inteligencia artificial ministra anticorrupción: Así es Diella», *Eleconomista.es*, sección Tecnología. 16/9/2025. Disponible en: <https://www.eleconomista.es/tecnologia/noticias/13548088/09/25/albania-nombra-a-una-inteligencia-artificial-ministra-anticorrupcion-asi-es-diella.html> (consultado el 18/9/2025).

³⁰ TRANSPARENCIA INTERNACIONAL. *Índice de percepción de la corrupción 2024* (IPC-2024). Disponible en: <https://transparencia.org.es/que-hacemos/investigacion/indices/indice-de-percepcion-de-la-corrupcion-2024/> (consultado el 18/9/2025).

teniendo en cuenta la velocidad tan desbordante a la que se está produciendo este salto tecnológico.

La opacidad, la pérdida de trazabilidad, los sesgos (involuntarios o no), las castas tecnológicas y el analfabetismo digital pueden generar una serie de vacíos legales que permitan la dilución de responsabilidades ante consecuencias negativas. Esto llevaría a una degradación paulatina del sistema, iniciando el temido proceso de «apatía cívica» por el cual el ciudadano pierde el interés por los procesos de toma de decisiones, incrementando la desconfianza y erosionando la legitimidad del contrato social establecido. Estos problemas son genéricos en cualquier forma de gobierno que ignore los riesgos que afronta, y la «algocracia» no parece ser la excepción.

Desarrollar unos marcos regulatorios eficientes, establecer sistemas de auditorías independientes y educar a la población para que comprendan estas nuevas tecnologías serían los pilares fundamentales para garantizar la transparencia. Todo ello enmarcado por el principio del «control humano significativo», de forma que siempre exista un criterio ético y moral con asunción de responsabilidades. Como hemos visto, esto es esencial en aquellas decisiones que pueden implicar daños a las personas, como ocurre en seguridad y defensa.

Automatización de documentación, predicción en infraestructuras, atención automatizada, recepción de denuncias, reconocimiento facial, evaluación de candidatos...; en las Administraciones públicas ya se están implementando sistemas «algocráticos» de forma efectiva. Algoritmos, redes neuronales, Big-Data o IA son conceptos que ya forman parte del lenguaje administrativo y, cada vez más, de los procesos de análisis y toma de decisiones de los Gobiernos. Por eso, se puede afirmar que ya existe un proceso de transformación progresivo, generando una «algocracia parcial» en la que todas nuestras decisiones estarán fundamentadas en información obtenida mediante estas tecnologías. El problema es decidir dónde estará el límite entre lo meramente consultivo o analítico y dónde permitir a los algoritmos tomar decisiones complejas.

En ambos extremos, el control humano debe ser inexcusable para evitar que estos sistemas alcancen la completa autonomía en sus decisiones («algocracia total»), estableciendo una jerarquía de responsables con capacidad para detener cualquier acción automática, pero atendiendo únicamente a la subjetividad del operador.

Pero, antes de llegar a ese punto, deberemos establecer si es viable aspirar a gobernarnos con el apoyo cada vez más creciente de sistemas de decisiones automatizadas o si eso supondrá sacrificar en exceso nuestra soberanía o nuestros derechos fundamentales. Al fin y al cabo, recordemos que el reto ya no es únicamente el desarrollo tecnológico, sino cómo asimilar esa nueva tecnología sin perder nuestra esencia como sociedad, nuestra perspectiva ética y el juicio moral humano.

*Alfredo Gil Laso**
Capitán CIP del ET, INTA