
Contestabilidad algorítmica en el Estado colombiano: un canal de objeción asistido por IA¹

Emely Condor & Federico Perez
(Perú) (Colombia)

Con
Apart Research

Resumen

Los sistemas algorítmicos del Estado colombiano ya toman o asisten decisiones que afectan derechos, pero los canales para objetar esas decisiones casi no existen. Construimos un Índice de Contestabilidad Ciudadana que operacionaliza, artículo por artículo, la Directiva Conjunta 007 de 2025 de la Procuraduría y la Defensoría, y lo aplicamos a 409 sistemas de decisión automatizada del repositorio de la Universidad de los Andes; sistemas de alto impacto se verificaron contra las webs oficiales de cada entidad. El resultado central es una brecha medible: la transparencia informacional es alta (las entidades publican qué es el sistema, su objetivo y sus datos) mientras la contestabilidad es casi nula. Solo 6 de 409 sistemas ofrecen un canal de objeción completo, 391 no ofrecen ninguno, y ninguno de los 154 sistemas de alto riesgo publica un análisis de impacto algorítmico. A partir de esta evidencia proponemos: el diseño de un protocolo de canal de objeción asistido por IA que recibe el reclamo de un ciudadano, lo clasifica por tipo de sistema y lo enruta al responsable según la Directiva 007. La conclusión para la seguridad de la IA es que la contestabilidad debe diseñarse antes de que los agentes IA lleguen, no después.

1. Introducción

En América Latina, la administración pública adopta sistemas algorítmicos a un ritmo mayor que la capacidad institucional para hacerlos rendir cuentas. En Colombia, un repositorio académico documenta 409 sistemas de decisión automatizada en uso por entidades estatales, de los cuales 265 (65%) emplean inteligencia artificial y 276 (67%) procesan datos personales. Estos sistemas focalizan subsidios, predicen riesgos de reincidencia o deserción, priorizan auditorías y atienden a poblaciones vulnerables. La pregunta que motiva este trabajo no es si el Estado debe usar estas herramientas, sino qué le ocurre a una persona cuando una de ellas decide en su contra.

El problema presente es de contestabilidad. Cuando un sistema excluye a alguien de un beneficio o lo clasifica como riesgo, el afectado necesita poder detectar que hubo una decisión automatizada, entenderla y objetarla ante un responsable. Por ejemplo, una persona que forma parte de una comunidad vulnerable puede ser excluida por un sistema de focalización por parte del gobierno de Colombia que, mediante IA, busca brindar un subsidio a esa comunidad, de modo que el subsidio termina afectándola. [La Directiva Conjunta 007 de 2025](#), expedida por la Procuraduría General de la Nación y la Defensoría del Pueblo en cumplimiento de la Sentencia T-067 de 2025 de la Corte Constitucional, convirtió esa exigencia en

¹ Investigación realizada en [The Global South AI Safety Hackathon](#), Junio 2026.

obligación. Todo sujeto obligado debe publicar un canal de objeción (art. 9h y 12) y, para los sistemas de mayor impacto, criterios de decisión y análisis de impacto algorítmico (art. 13).

El problema futuro es de seguridad de la IA. Los canales de objeción actuales, donde existen, presuponen un funcionario humano detrás de la decisión. Cuando agentes de IA empiecen a ejecutar de forma autónoma las funciones que hoy desempeñan funcionarios, la contestabilidad se vuelve estructuralmente imposible si no se diseña desde ahora, no habrá un humano responsable a quien apelar, el sistema tenderá a auto-justificar su propia decisión, y los canales actuales no fueron construidos para ese escenario. La ventana para incorporar la contestabilidad como propiedad de diseño se cierra a medida que la autonomía aumenta.

Nuestra contribución principal:

1. El diseño de un protocolo de canal de objeción asistido por IA que recibe un reclamo ciudadano, lo clasifica por el tipo de sistema (causal del art. 13) y lo enruta al responsable correcto según la Directiva 007.

Esta contribución se basa en que no queremos que ningún ciudadano se quede sin la oportunidad de solicitar un reclamo, y que este reclamo no se vea después de un año, cuando el daño ya haya incrementado, sino que pueda ser respondido de manera rápida y efectiva. Por ello, también creamos el ***Índice de Contestabilidad Ciudadana*** para operacionalizar la Directiva 007 variable por variable y lo aplicamos a los 409 sistemas, con casos de alto impacto verificados contra fuente primaria. El índice cuantifica, por primera vez para Colombia, la brecha entre transparencia informacional (alta) y contestabilidad (casi nula). Asimismo, a manera complementaria, hemos propuesto una recomendación de política pública ubicada en el Apéndice A, sobre una **cláusula de contestabilidad heredada para la era de los agentes de IA**.

2. Trabajo relacionado

Nuestro punto de partida normativo es la Sentencia T-067 de 2025, que reconoció la transparencia algorítmica, en sus facetas activa y pasiva, como parte del núcleo esencial del derecho de acceso a la información pública, y la Directiva Conjunta 007 que la desarrolla.

En el plano de la medición, el *Repositorio Sistemas automatizados de toma de decisiones en el sector público de Colombia* de la Universidad de los Andes, mapea la existencia de sistemas algorítmicos del Estado. Se utilizó las bases de datos para poder medir, sistema por sistema, si el ciudadano puede objetar una decisión. El informe de GPAI (2025) sobre transparencia algorítmica en el sector público y el trabajo de Garrido, Lapostol y Hermosilla (2021) proveen el marco conceptual de transparencia activa y pasiva que adoptamos. Nuestro *Índice de Contestabilidad Ciudadana* evalúa si esa información habilita efectivamente la objeción ciudadana.

En el plano del diseño, nos apoyamos en la literatura de *contestable AI by design* (Alfrink et al., 2023; Almada, 2019), que argumenta que la posibilidad de impugnar una decisión automatizada debe construirse en el sistema y no añadirse después, y en la tradición del artículo 22 del RGPD europeo y la Directiva canadiense sobre decisiones automatizadas, ambas citadas por la propia Directiva 007. Nuestra cláusula de contestabilidad extiende ese principio al caso aún no tratado en esa literatura, la sustitución del decisor humano por un agente autónomo, donde el riesgo no es solo una mala explicación sino la desaparición del destinatario de la apelación. El trabajo de Derechos Digitales (2022) sobre la judicialización del acceso a beneficios sociales documenta empíricamente la consecuencia de ese vacío en la región y motiva la urgencia.

3. Métodos

3.1 Universo e instrumento

Partimos de los 409 sistemas de decisión automatizada del repositorio de la Universidad de los Andes (marzo de 2026), que incluye tanto sistemas de IA como de automatización robótica de procesos, en línea con el alcance del artículo 6 de la Directiva.

El índice tiene dos capas que reflejan la estructura de la Directiva. La capa base (11 variables, B1 a B11) se ancla a los artículos 9 y 12 y aplica a todo sistema: existencia y nombre, objetivo, función institucional, estado, declaración de datos personales, tipo de datos, desarrollador, canal de objeción, contacto para información, actualización y lenguaje claro. La capa de alto riesgo (4 variables, H1 a H4) se ancla al artículo 13 y aplica solo a los sistemas que asignan o niegan beneficios, predicen riesgos, perfilan personas, infieren emociones, atienden poblaciones vulnerables o se relacionan con la salud: criterios de decisión, justificación del uso, análisis de impacto algorítmico y medidas de mitigación. Cada variable se codifica como 0, 0.5 o 1 según una rúbrica explícita; el subíndice de cada capa es el promedio ponderado de sus variables, y el índice final pondera la capa base en 0.6 y la de alto riesgo en 0.4 para los sistemas de alto riesgo, y equivale a la capa base para el resto. Lo que no está publicado no es contestable, de modo que la ausencia de información se codifica como incumplimiento, en coherencia con el principio de máxima divulgación del artículo 5 de la Directiva. La variable de canal de objeción (B8) recibe el mayor peso dentro de la capa base, porque es la que materializa el derecho a impugnar.

3.2 Clasificación de alto riesgo, verificación y enrutamiento

La clasificación de alto riesgo se construyó con una heurística reproducible de palabras clave y sector sobre los campos de nombre, resumen, objetivo, descripción, sector y población, contrastada con verificación humana, mapeando cada sistema a los siete supuestos del artículo 13. La puntuación de las variables se realizó de forma automatizada a partir de los campos del repositorio, y sistemas de alto impacto que asignan o niegan beneficios se recodificaron tras consultar las webs oficiales de cada entidad y la jurisprudencia asociada. Para el triaje del prototipo añadimos al instrumento tres variables del repositorio: interacción con el ciudadano (G2C), uso de datos personales y estado del sistema.

3.3 Protocolo de canal de objeción asistido por IA que recibe el reclamo de un ciudadano.

Nuestra propuesta sobre el protocolo de canal de objeción (ver Figura 1) se diseña como una capa de enrutamiento sobre este instrumento mencionado. Dado un reclamo, el sistema (i) identifica el sistema algorítmico aludido, (ii) lo clasifica por su causal del artículo 13, y (iii) consulta la variable de canal de objeción: si el sistema tiene un canal real, enruta a él; si no, escala al contacto institucional de la entidad responsable como destinatario por defecto, dejando registro del vacío. La priorización usa el núcleo crítico definido como la intersección de alto riesgo, uso de datos personales e interacción directa con el ciudadano.

Canal de Objeción Asistido por IA — Arquitectura de Enrutamiento

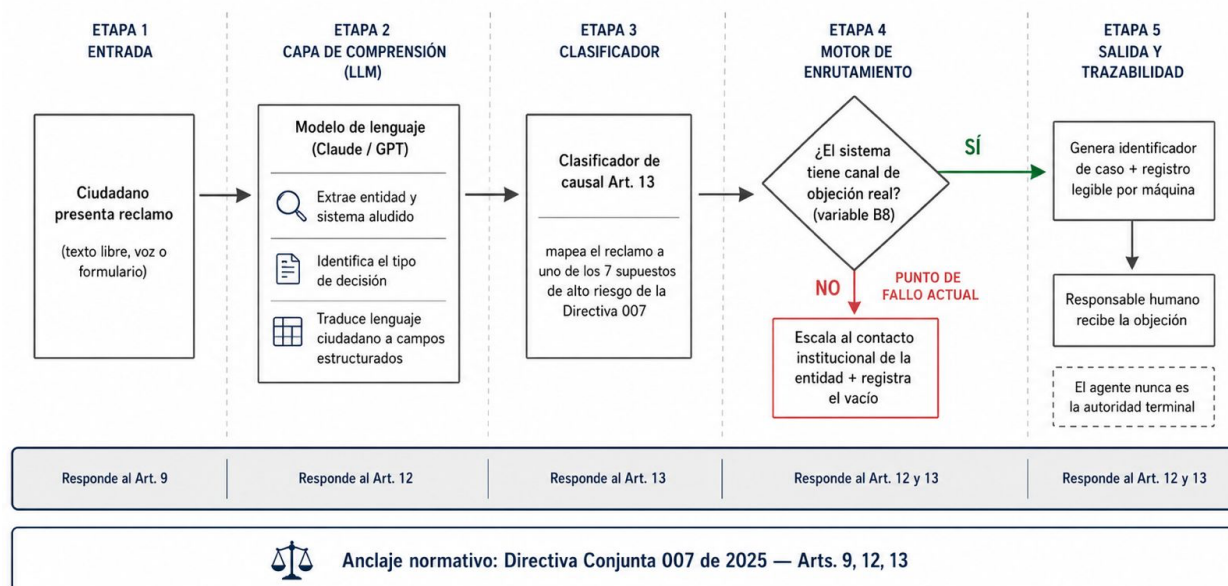


Figura 1. Propuesta del diseño de un protocolo de canal de objeción asistido por IA que recibe el reclamo de un ciudadano.

4. Resultados

Dentro de los hallazgos centrales, existe una brecha medible entre dos transparencias. Las entidades publican bien qué es cada sistema y qué hace, pero casi nunca cómo objetarlo. La Tabla 1 resume los indicadores del universo completo tras la verificación.

Tabla 1. Indicadores de contestabilidad, 409 sistemas (índice normalizado, pesos a 100%).

Indicador	Valor
Sistemas de decisión automatizada documentados	409
Usan inteligencia artificial / usan RPA	265 / 144
Usan datos personales	276 (67%)
De cara al ciudadano (G2C)	160 (39%)
Sistemas de alto riesgo (art. 13)	154 (38%)
Núcleo crítico (alto riesgo + datos personales + G2C)	13
Con canal de objeción completo (B8 = 1)	6
Con algún canal de objeción (B8 > 0)	18
Sin ningún canal de objeción (B8 = 0)	391 (96%)
Sistemas de alto riesgo con análisis de impacto (H3 = 1)	0 de 154
Índice promedio: global / alto riesgo / base	0.45 / 0.36 / 0.51

La Figura 2 desagrega el índice por variable y hace visible la brecha. Las variables descriptivas puntúan alto (objetivo 0.99, estado 1.00, lenguaje claro 0.99, desarrollador 0.85), mientras que la variable de canal de objeción cae (0.03) y el análisis de impacto de los sistemas de alto riesgo es cero en el 100% de los casos. La transparencia que el Estado entrega es la que menos cuesta, mientras la que habilitaría al ciudadano a defenderse no se entrega.

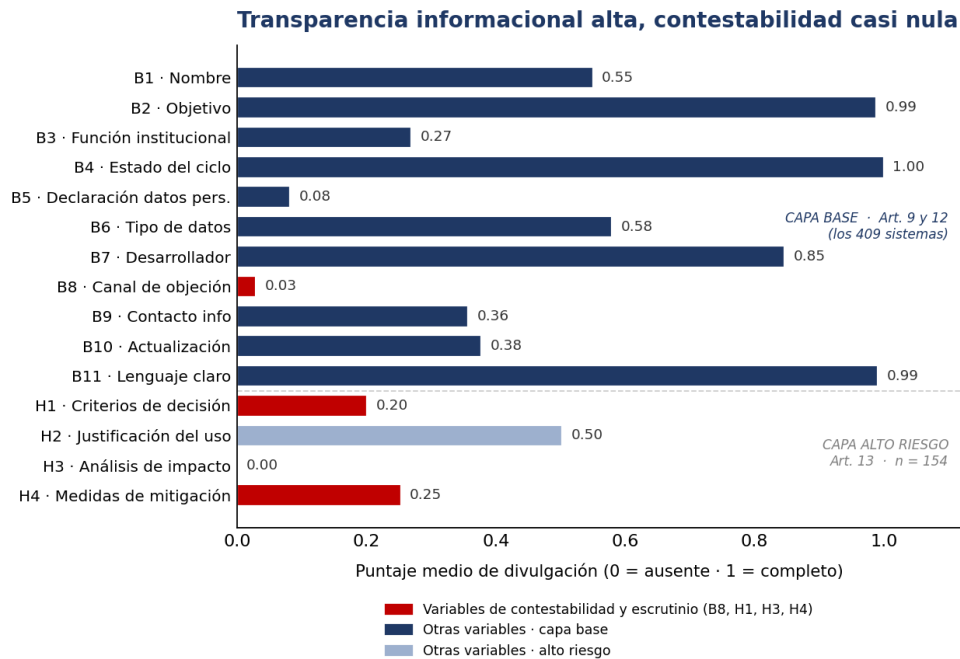


Figura 2. Puntaje medio de divulgación por variable (0 = ausente, 1 = completo). En rojo, las variables de contestabilidad y escrutinio.

La verificación contra fuente primaria confirma que el patrón no es un artefacto de la codificación automática. La Tabla 2 muestra los seis asignadores de beneficios verificados. En todos, el canal de objeción específico a la decisión automatizada está ausente o es apenas un buzón genérico de peticiones, y ninguno publica análisis de impacto.

Tabla 2. Ejemplos de sistemas de alto impacto del Estado colombiano que asignan o niegan beneficios y servicios (art. 13.i) o predicen riesgos sobre las personas (art. 13.ii)

ID	Sistema (entidad)	Función	Canal objeción	Anál. impacto	Recurso real
5 y 6	Equidad Digital (DPS)	Focalización de programas sociales	Ausente	No	Tutela
4	Antifraudes (DPS)	Detección de fraude en beneficios	Ausente	No	Tutela
2	BETTO (ICBF)	Focalización de beneficiarios	Genérico	No	PQRS
10	Modelo de vulneración NNA (ICBF)	Riesgo de vulneración de menores	Genérico	No	PQRS
406	Asignación de sede (ICFES)	Asignación de lugar de examen	Genérico	No	PQRS

Nota: Equidad Digital comprende dos sistemas (ID 5 y 6); las demás filas corresponden a un sistema cada una.

El caso de la focalización del DPS es ilustrativo. En la Sentencia T-312 de 2021, un ciudadano excluido del listado de beneficiarios de Ingreso Solidario solo pudo objetar la decisión mediante una acción de tutela, porque no existía un canal administrativo para hacerlo. La tutela es hoy el canal de facto: lento, costoso y condicionado a que la persona sepa que un algoritmo decidió. Nuestro prototipo es precisamente el escalón administrativo ausente entre la decisión y el juez.

5. Discusión y limitaciones

La contestabilidad es una propiedad de seguridad, es el mecanismo por el cual los humanos conservan control sobre decisiones automatizadas que los afectan. Nuestra evidencia muestra que esa propiedad ya está ausente con funcionarios humanos todavía en el circuito. El escenario que se aproxima, en el que agentes de IA ejecuten estas funciones de forma autónoma, agrava el problema por tres vías: desaparece el humano responsable a quien apelar, el propio sistema tiende a generar la justificación de su decisión (autojustificación), y los canales actuales no fueron diseñados para recibir objeciones contra un agente. Si la contestabilidad no se construye antes de esa transición, se vuelve estructuralmente irrecuperable.

Limitaciones

La limitación principal consideramos que es temporal. El repositorio de la Universidad de los Andes está actualizado hasta marzo de 2026, mientras que la Directiva 007 entró en vigor apenas unos seis meses antes (septiembre 2025), de modo que las entidades han tenido poco tiempo para adaptar sus prácticas de publicación a las nuevas obligaciones. Nuestra medición captura, en consecuencia, un momento temprano del cumplimiento, cuando muchas entidades aún no han implementado lo que la norma exige. Esto hace que el estudio funcione como una línea base valiosa, así replicarlo más adelante permitiría medir si las entidades efectivamente cierran la brecha a medida que avanza la adopción de la Directiva, y distinguir el rezago inicial de un incumplimiento persistente.

Trabajo futuro

- El principal desarrollo posterior se desprende del carácter de línea base de la medición. Al coincidir con la entrada en vigor de la Directiva 007, el índice puede aplicarse de nuevo a intervalos regulares para evaluar si la regulación de transparencia algorítmica produce cambios efectivos en la conducta de las entidades. Un seguimiento longitudinal del índice respondería esa pregunta y aportaría evidencia cuantitativa a la mesa de seguimiento prevista en la propia Directiva.
- Implementación del canal de objeción. El protocolo descrito en este trabajo debe construirse como un sistema funcional, capaz de recibir reclamos reales, clasificarlos por la causal del artículo 13 y enrutarlos al responsable correspondiente, y ponerse a prueba en un piloto con una entidad pública. Esta fase permitiría medir la precisión del enrutamiento, depurar el clasificador con casos reales y documentar el efecto del canal sobre el tiempo y el costo de objetar una decisión, hoy mediados casi siempre por la acción de tutela.
- En el plano metodológico, ampliar la verificación manual a una muestra estratificada por sector permitiría calibrar la codificación automática y precisar los puntajes del índice; y una prueba con usuarios simulados sobre los pocos canales existentes mediría si funcionan de hecho, y no solo si están publicados.
- Formalizar la cláusula de contestabilidad heredada como modelo contractual, haría de la transición hacia los sistemas de agentes un objeto regulable desde el contrato y no solo un argumento conceptual.

- Nuestra propuesta del diseño de un protocolo de canal de objeción asistido por IA (ver Figura 1) es escalable a la región, y el caso colombiano lo hace especialmente relevante. En Colombia, un país avanzado en temas de IA, que la contestabilidad sea casi nula incluso donde ya existe un marco legal es una señal de alerta para toda la región. Aplicar nuestro *Índice de Contestabilidad Ciudadana*, ajustando el anclaje a la regulación local, permitiría confirmar esta hipótesis y convertir la contestabilidad en un criterio comparable de gobernanza algorítmica en América Latina.

6. Conclusión

Medimos la contestabilidad de 409 sistemas algorítmicos del Estado colombiano frente a la Directiva 007 y encontramos que la transparencia sobre los sistemas es alta. Sin embargo, los canales para objetar son casi inexistentes, solo 6 de 409 ofrecen uno completo. Esta ausencia ya perjudica a los ciudadanos hoy, y empuja la única vía de objeción hacia los tribunales.

La contribución de gobernanza es tratar la contestabilidad como una propiedad que debe diseñarse antes de que los agentes de IA reemplacen a los funcionarios. El diseño de protocolo de canal de objeción cierra la brecha presente sobre la infraestructura normativa que ya existe, y nuestra propuesta adicional de política pública sobre la cláusula de contestabilidad heredada impide que esa brecha se vuelva permanente cuando desaparezca el responsable humano.

Código y datos

- **Datos e instrumento:** [Índice de Contestabilidad Ciudadana](#).
- **Prototipo:** diseño del protocolo descrito en la sección 3.3.

Referencias

- [1] Corte Constitucional de Colombia (2025). Sentencia T-067 de 2025.
- [2] Procuraduría General de la Nación y Defensoría del Pueblo (2025). Directiva Conjunta No. 007: estándares sobre transparencia algorítmica para los sistemas algorítmicos utilizados por el Estado. 30 de septiembre de 2025.
- [3] Congreso de la República de Colombia (2014). Ley 1712 de 2014, Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional.
- [4] Corte Constitucional de Colombia (2021). Sentencia T-312 de 2021.
- [5] Universidad de los Andes (2026). Repositorio Sistemas automatizados de toma de decisiones en el sector público de Colombia. Versión 3.
- [6] Global Partnership on Artificial Intelligence- GPAI (2025). Algorithmic transparency in the public sector: a report on the state of the art of algorithmic transparency instruments.
- [7] Garrido, R., Lapostol, J. P. y Hermosilla, M. P. (2021). Transparencia algorítmica en el sector público. GobLab UAI y Consejo para la Transparencia.
- [8] OECD (2019, rev. 2024). Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449.
- [9] Government of Canada (2019, act.). Directive on Automated Decision-Making.
- [10] Parlamento Europeo y Consejo (2016). Reglamento General de Protección de Datos (RGPD), artículo 22.
- [11] Alfrink, K., Keller, I., Kortuem, G. y Doorn, N. (2023). Contestable AI by Design: Towards a Framework. Minds and Machines.

- [12] Almada, M. (2019). Human intervention in automated decision-making: toward the construction of contestable systems. Proceedings of ICAIL.
- [13] Derechos Digitales (2022). Inteligencia artificial en la función pública: judicialización del acceso a beneficios sociales y datos sucios.
- [14] Guía Ética para la Implementación, Desarrollo y Uso de Sistemas de Inteligencia Artificial en Entidades Públicas de Colombia (2026). Gobierno de Colombia.

Apéndice A. Recomendación de política: la cláusula de contestabilidad heredada para la era de los agentes de IA

Este apéndice formula una recomendación de política orientada al escenario en que agentes de IA empiecen a ejecutar de forma autónoma funciones hoy desempeñadas por funcionarios públicos, quienes ya toman decisiones sobre sistemas de alto riesgo. La presentamos como propuesta de gobernanza.

El principio es que la obligación de contestabilidad debe ir con la función. Si una función pública la desempeñaba una persona con un canal de objeción asociado, cualquier agente que la reemplace hereda el deber de mantener ese canal operativo. Sin esta regla, la sustitución del decisor humano por un sistema autónomo elimina al destinatario mismo de la apelación, y con él la posibilidad de impugnar.

Como requisito mínimo, todo agente IA que sustituya la decisión de un funcionario debería cumplir cinco condiciones:

- i. emitir un registro de decisión legible por máquina con la decisión, los insumos y los criterios invocados;
- ii. exponer un punto de contestación que reciba objeciones y devuelva un identificador de caso;
- iii. designar un responsable humano que reciba las objeciones escaladas, de modo que el agente nunca sea la autoridad terminal;
- iv. generar la explicación al ciudadano a partir del registro de decisión;
- v. registrarse, con su punto de contestación, en el reporte anual de la Directiva.

Esta cláusula es implementable hoy por la Directiva 007, que ordena incluir cláusulas de transparencia en la contratación de sistemas algorítmicos. Formalizarla como modelo de cláusula contractual permitiría que cada nueva adquisición de un agente IA quede obligada, desde el contrato, a preservar el canal de objeción, cerrando la ventana antes de que la transición a la autonomía la vuelva irrecuperable.

Declaración de uso de LLM

Se emplearon herramientas de inteligencia artificial como apoyo en tareas acotadas. Claude (Anthropic) asistió en el procesamiento de los datos para el cálculo del índice y en la redacción y edición del texto, y ChatGPT (OpenAI) se utilizó para elaborar el diagrama de arquitectura de la Figura 1. Las codificaciones se contrastaron contra fuentes primarias y la totalidad de los datos y afirmaciones fue revisada por el equipo. La versión final del documento es responsabilidad de los autores.