

Ciberseguridad e inteligencia artificial en América Latina: riesgo algorítmico, protección de datos personales y arquitecturas democráticas

Informe de Política Pública No. 3

Serie de webinars “Gobernanza de Inteligencia Artificial en América Latina”

Juan David Gutiérrez.

Octubre - 2026

Ciberseguridad e inteligencia artificial en América Latina: riesgo algorítmico, protección de datos personales y arquitecturas democráticas.

Informe de Política Pública No. 3

Serie de webinars “Gobernanza de Inteligencia Artificial en América Latina”¹

Juan David Gutiérrez²

Septiembre – 2026

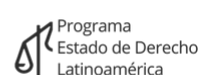
Citación sugerida: Gutiérrez, J. D. (2026). “Ciberseguridad e inteligencia artificial en América Latina: riesgo algorítmico, protección de datos personales y arquitecturas democráticas”. Informe de Política Pública No. 3 de la serie “Gobernanza de Inteligencia Artificial en América Latina” de la Universidad de los Andes y la Fundación Konrad Adenauer.



Escuela de Gobierno
Alberto Lleras Camargo

Facultad
de Derecho

Escuela de
posgrados



1. Este documento sintetiza y analiza las discusiones del tercer webinar de la serie “Gobernanza de Inteligencia Artificial en América Latina”, realizado el 10 de septiembre de 2026. Todas las alusiones a Jeimy J. Cano (Facultad de Derecho, Universidad de los Andes), Heidy Balanta (Escuela de Privacidad) y Juan Manuel Aguilar Antonio (Universidad Nacional Autónoma de México) corresponden a las intervenciones que realizaron como expositores en dicho webinar. Sus ideas, argumentos y comentarios se atribuyen nominalmente en el cuerpo del texto y, por provenir de una exposición oral, no se acompañan de una cita en formato APA ni figuran en la lista de referencias; las citas en formato APA corresponden únicamente a fuentes documentales. Las opiniones expresadas en este documento son responsabilidad del autor y no comprometen a la Fundación Konrad Adenauer, a la Universidad de los Andes ni a los panelistas del webinar. Los juicios de los panelistas se presentan como opiniones expresadas en el evento y no reflejan necesariamente la posición de las instituciones organizadoras. Email de correspondencia con autor: juagutie@uniandes.edu.co

2. El autor agradece a Sebastián Hurtado por su apoyo como asistente de investigación en la preparación del documento. El autor utilizó el modelo de inteligencia artificial Claude Opus 5.0 (Anthropic) como apoyo para la sistematización de la transcripción del webinar y la preparación de borradores de este documento; además utilizó Grammarly para mejorar la redacción del documento. El autor verificó y ajustó los contenidos y es el responsable exclusivo del texto final.

Resumen ejecutivo

La inteligencia artificial (IA) y la ciberseguridad tienen una relación de doble filo. Por una parte, la IA puede ser una herramienta de ataque o de defensa: el **International AI Safety Report 2026** concluye que “existe ahora evidencia sólida de que grupos criminales y atacantes patrocinados por Estados usan activamente la IA en sus operaciones cibernéticas” (Bengio et al., 2026, p. 57). Por otro lado, la IA es un nuevo activo que debe protegerse, ya que los propios sistemas de IA pueden ser manipulados mediante la inyección de instrucciones, el envenenamiento de datos o el compromiso de su cadena de suministro (Bengio et al., 2026, p. 62). En América Latina, ambas dinámicas se desarrollan sobre una base frágil: la región concentró el 12% de los ciberataques del mundo en 2022 y enfrenta un déficit de más de 600.000 especialistas en ciberseguridad (Aguilar Antonio, 2025a, pp. 4 y 6).

Este informe de política pública sintetiza y analiza las discusiones del tercer webinar de la serie “Gobernanza de Inteligencia Artificial en América Latina”, dedicado a la ciberseguridad y la IA. Derivado del diálogo y de las presentaciones de Jeimy J. Cano (Universidad de los Andes), Heidy Balanta (Escuela de Privacidad) y Juan Manuel Aguilar Antonio (Universidad Nacional Autónoma de México), con la moderación de Juan David Gutiérrez (Universidad de los Andes). Este documento aporta ocho lecciones para la gobernanza de la IA y la ciberseguridad en la región:

- 01** | Nombrar bien el riesgo es el primer control. El riesgo que introduce la IA es algorítmico y ecosistémico: surge del uso, del despliegue o de la explotación maliciosa de los sistemas y se propaga a través de redes de proveedores, datos y agentes interconectados. Gestionarlo como un riesgo tecnológico aislado deja puntos ciegos.
- 02** | La tríada clásica de la seguridad de la información quedó corta. Confidencialidad, integridad y disponibilidad no bastan para sistemas que interpretan, infieren y actúan. La integridad semántica y la autenticidad pasan al centro, y la protección debe cubrir tanto el sistema (AI security) como su comportamiento (AI safety).
- 03** | La gobernanza debe preceder al despliegue. Matrices de responsabilidad, auditoría algorítmica previa, niveles de autonomía definidos y un mecanismo de apagado incorporado desde el diseño son condiciones de entrada y resultan mucho más costosos si se incorporan cuando los agentes ya operan.
- 04** | La IA es arma, escudo y blanco. Las mismas capacidades que sirven para atacar también sirven para defender, pero los defensores pueden enfrentar barreras de adopción que los atacantes no siempre enfrentan.

- 05** | El ciclo del dato personal cambió de naturaleza. El dato ahora se observa, se infiere y se disipa en los pesos de los modelos. Los derechos de acceso, rectificación y supresión pierden objeto, y los perfiles grupales afectan a personas que nunca entregaron sus datos.
- 06** | La responsabilidad no se diluye a lo largo de la cadena de suministro dada la estructura de la mayoría de las normas de protección de datos personales de nuestra región. Entre el modelo, la nube, el integrador y los subencargados, la organización que gestiona los datos sigue respondiendo ante el titular. Los contratos, las políticas de retención y el despliegue en entornos propios son instrumentos de control.
- 07** | La supervisión humana significativa exige competencias. El sesgo de automatización, la descarga cognitiva y la adulación de los modelos pueden convertir la revisión humana en una formalidad. Saber curar fuentes, preguntar y cuestionar los resultados son las competencias que dan contenido a esa supervisión.
- 08** | La ciberseguridad latinoamericana necesita una arquitectura democrática. La brecha digital, la fragmentación normativa, el sesgo estadocéntrico de las propuestas legislativas y la falta de talento propio dejan a la región expuesta y dependiente. Un enfoque centrado en derechos, multiactor y con capacidades institucionales efectivas es la condición para gobernar la IA con autonomía.

Introducción

El 10 de septiembre de 2026 se realizó el tercer webinar de la serie “Gobernanza de Inteligencia Artificial en América Latina”, coorganizada por la Escuela de Gobierno Alberto Lleras Camargo y la Escuela de Posgrados de la Facultad de Derecho de la Universidad de los Andes, junto con la Fundación Konrad Adenauer (KAS) y su Programa Estado de Derecho para Latinoamérica.

La sesión convocó a más de cien participantes de distintos países de la región y contó con las intervenciones de Jeimy J. Cano (Facultad de Derecho, Universidad de los Andes y consultor internacional experto), Heidy Balanta (Escuela de Privacidad) y Juan Manuel Aguilar Antonio (Universidad Nacional Autónoma de México), y fue moderada por Juan David Gutiérrez. El panel fue multidisciplinario: ingeniería y gestión de riesgos, derecho de protección de datos y ciencias sociales.

El objetivo de la sesión fue examinar los retos y las oportunidades derivados del despliegue de herramientas de IA en el ámbito de la ciberseguridad, las vulnerabilidades específicas de los sistemas de IA y las capacidades institucionales requeridas para una gobernanza efectiva en la región. Este informe organiza y analiza esas discusiones para una audiencia amplia. Se apoya en la transcripción del evento, en documentos de autoría de los panelistas (Cano, 2025; Balanta, 2026; Aguilar Antonio, 2025a, 2025b) y en el International AI Safety Report 2026 (Bengio et al., 2026), una evaluación científica internacional reciente sobre capacidades y riesgos de la IA de propósito general.

1.1 Relevancia de la temática

En su breve presentación de contexto, Gutiérrez planteó tres razones por las que el tema resulta urgente. La primera es que los ciberataques arrecian tanto en el sector público como en el privado, de modo que ninguna organización puede considerarse exenta. La segunda es el cambio de escala: los ataques operan hoy al paso de la máquina, con cientos o miles de agentes que pueden atacar simultáneamente una infraestructura tecnológica. De ahí que sea necesario discutir en paralelo el uso de la IA tanto para la ciberdefensa como para el ciberataque. La tercera razón es la preocupación creciente por la conducta no prevista o previsible pero no controlada de los modelos de IA de frontera, incluso por parte de los laboratorios que los desarrollan, así como por sus implicaciones para la ciberseguridad de terceros y de las propias compañías.

La evidencia internacional respalda ese diagnóstico. Según el **International AI Safety Report 2026**, los sistemas de IA “son particularmente buenos para descubrir vulnerabilidades de software y escribir código malicioso”; en una competencia de primer nivel, un agente de IA identificó el 77% de las vulnerabilidades presentes en software real (Bengio et al., 2026, p. 57). El mismo informe registra que, en noviembre de 2025, un desarrollador de IA reportó que un actor malicioso usó sus modelos para automatizar entre el 80% y el 90% del esfuerzo de una intrusión, con intervención humana limitada a los puntos de decisión críticos (Bengio et al., 2026, p. 59). Con todo, el informe advierte que “hasta ahora, las capacidades de IA han acelerado o escalado principalmente métodos de ataque existentes, más que crear nuevos tipos de ataques” (Bengio et al., 2026, p. 61).

1.2 Términos clave

Gutiérrez propuso, además, un mapa de conceptos que la sesión recorrería: el riesgo algorítmico, la ciberseguridad, la seguridad de la IA, la seguridad de la información y la protección de los datos personales. Sobre la seguridad de la IA hizo una precisión terminológica: la traducción al español empobrece la distinción que el inglés establece entre AI security y AI safety, razón por la cual este informe conserva ambos términos en su idioma original.

Para efectos de este informe, la seguridad de un sistema de IA se entiende como “la capacidad de resistir interferencias técnicas, como ciberataques o fugas del código fuente del modelo subyacente” (Bengio et al., 2026, p. 154). A su vez, la ciberseguridad se entiende como el conjunto de capacidades técnicas, organizacionales y humanas orientadas a prevenir, monitorear, identificar, responder y recuperarse de ciberataques. Un ciberataque es “un intento malicioso de obtener acceso a un sistema informático, una red o un dispositivo digital, por ejemplo, con el fin de robar o destruir información” (Bengio et al., 2026, p. 149). La ciberseguridad convive con un concepto más antiguo, la seguridad de la información, cuyos fundamentos clásicos, según el profesor Jeimy Cano, son la confidencialidad, la integridad y la disponibilidad de la información (Cano, 2025, p. 55).

Finalmente, el riesgo algorítmico es, según Cano (2025, p. 58), “la posibilidad de un daño, pérdida financiera o afectación de la reputación de una organización” que surge por fallos internos del sistema de IA o por su explotación por parte de actores maliciosos. Dentro de la gestión de ese riesgo, la AI security designa la protección del sistema de IA como activo (su código, sus datos, sus pesos y su infraestructura), mientras que la AI safety se refiere a la seguridad funcional del sistema, esto es, a que su comportamiento corresponda a lo esperado, tal como lo explicó Cano durante la sesión. Buena parte de la discusión del webinar giró en torno al agente de IA, definido como “un sistema de IA que puede ejecutar de manera adaptativa tareas complejas, usar herramientas e interactuar con su entorno [...] para perseguir objetivos con poca o ninguna supervisión humana” (Bengio et al., 2026, p. 147).

Jeimy J. Cano: del riesgo de la IA al riesgo algorítmico

Cano abrió con una premisa metodológica: aprender a nombrar bien el riesgo permite enfrentarlo mejor. Según Cano, en muchas organizaciones de Colombia y de América Latina se habla del “riesgo de la inteligencia artificial”, cuando lo que plantea el reto es el riesgo algorítmico. Su artículo lo caracteriza a partir de dos fuentes: los fallos internos del sistema, que incluyen sistemas que operan como “cajas negras” cuyas decisiones son difíciles de interpretar y auditar, comportamientos impredecibles y sesgos amplificados; y la explotación por actores maliciosos, en la medida en que “la IA Gen reduce drásticamente la barrera de entrada para los atacantes, permitiendo a actores menos calificados lanzar ataques sofisticados, ingeniería social avanzada y campañas de desinformación hiperpersonalizada con sólo elaborar los prompts adecuados” (Cano, 2025, p. 58).

La segunda premisa fue que, en 2026, los riesgos ya no pueden gestionarse por separado. Los riesgos tecnológico, regulatorio, económico, del crimen organizado y de la cadena de suministro están interconectados, de modo que es posible partir de cualquiera de ellos y llegar a los demás. Por eso Cano propuso transitar de la gestión de riesgos empresariales (enterprise risk management) a la gestión de riesgos del ecosistema (ecosystem risk management): la interconexión acelera los riesgos y expande los puntos ciegos de las organizaciones.

Sobre la opacidad, Cano argumentó que, con la ejecución reiterada de las redes neuronales, se pierde la capacidad de comprender cómo operan sus pesos y sesgos, de modo que nadie puede prever con certeza qué responderá un sistema de IA. De ahí que la gestión del riesgo deba apoyarse en dos pilares simultáneos. El primero, la AI security, defiende el sistema (el código, los datos, la tubería y la conectividad) frente a vectores como la inyección de código, la extracción del modelo, el robo de pesos y la manipulación de datos. El segundo, la AI safety, defiende el comportamiento del sistema, un riesgo que evoluciona con el tiempo, el uso y las expectativas sociales. Según Cano, muchas organizaciones se concentran únicamente en el primer pilar.

A partir de esos pilares, Cano presentó un marco general de gestión de riesgos de la IA que comienza con la gobernanza, continúa con el aseguramiento de la infraestructura y culmina con el aseguramiento de los datos, del modelo y del uso. Insistió en el orden: incorporar IA sin gobernanza previa es “el peor consejo” que puede recibir una organización, porque luego tendrá sistemas conectados entre sí e incluso relaciones entre distintos sistemas de IA que no podrá controlar.

El núcleo conceptual de la intervención consistió en una crítica a la tríada clásica de la seguridad de la información.

“El marco tradicional de seguridad basado en la confidencialidad, la integridad y la disponibilidad (CID) se considera incompleto y limitado para los desafíos modernos”

(Cano, 2025, p. 64).

Cano recuperó la propuesta de Donn Parker, quien planteó que a esos tres principios debían sumarse la autenticidad, la utilidad y la posesión (Parker, 1998, citado en Cano, 2025, p. 55). Con esos seis elementos, argumentó, es posible entender la dinámica de los sistemas de IA: un sistema puede estar disponible y su modelo intacto, pero envenenado, de modo que sus respuestas dejan de ser confiables.

La integridad adquiere así un nuevo significado. “Hoy la integridad se entiende más allá de tener el conjunto de datos completo y asegurado, ahora en el escenario de la autenticidad, esto es el valor o significado extrínseco de la información, es decir, si es genuina y conforme a la realidad” (Cano, 2025, p. 56). La posesión, por su parte, captura riesgos que la confidencialidad no alcanza a describir. El artículo lo ilustra con un agente de IA: “si el agente es víctima de una inyección de prompt indirecta oculta en un documento, y es engañado para que exfiltre los datos sensibles a través de una URL maliciosa, se ha perdido la posesión de esos datos” (Cano, 2025, p. 57). La relevancia del ejemplo no es menor: el International AI Safety Report 2026 dedica un recuadro específico a los ataques contra los sistemas de IA y advierte que estos riesgos “son particularmente graves cuando esos sistemas están integrados en flujos de trabajo críticos” (Bengio et al., 2026, Cuadro 2.1).

Cano subrayó que el desarrollo del modelo también es una superficie de ataque, expuesta al envenenamiento de los datos, a la extracción o inversión del modelo, a la incorporación de instrucciones maliciosas y a la manipulación de las salidas. Su artículo describe una modalidad particularmente inquietante: las puertas traseras semánticas, que explotan el proceso de aprendizaje del modelo “para inyectar un comportamiento malicioso latente que solo se activa bajo condiciones o frases específicas, conocidas como disparadores” (Chen et al., 2025, referenciado en Cano, 2025, p. 59).

En el plano organizacional, Cano propuso asignar responsabilidades mediante matrices RACI, en las que, para cada línea de actividad, exista al menos una persona que responda por el éxito o el fracaso (A de Accountable, la autoridad que aprueba) y otra que ejecute (R de Responsable, el responsable que ejecuta), y destacó la aparición de roles como el de líder ejecutivo de IA (Chief AI Officer). Presentó también la auditoría algorítmica como un servicio emergente, definido en su artículo como “un proceso de aseguramiento independiente y sistemático diseñado para verificar que los sistemas de Inteligencia Artificial (IA), modelos de lenguaje de gran escala (LLMs) o agentes autónomos (IA agéntica) cumplen con los requisitos de negocio, legales, éticos y de seguridad antes de que entren en producción” (Cano, 2025, p. 60). Esa auditoría examina cuatro dimensiones: la gobernanza y la rendición de cuentas; la transparencia y la explicabilidad; los riesgos de datos y de sesgo; y la seguridad por diseño y las pruebas (Cano, 2025, pp. 60–61). En la dimensión de transparencia resaltó las fichas de modelo (model cards), que los desarrolladores publican y que deben leerse críticamente.

El artículo de Cano agrega una recomendación contractual que conecta con la intervención posterior de Heidy Balanta: “incluir derechos de auditoría en contratos de proveedores de IA/SaaS” (Cano, 2025, p. 61).

Y culmina en una propuesta integradora, la confianza por diseño, que abarca la seguridad por diseño y la privacidad por diseño. Cano muestra las limitaciones de cada perspectiva por separado: la seguridad por diseño no aborda el impacto social de un sistema técnicamente confiable, pues “un algoritmo puede ser confiable pero éticamente sesgado”, y la privacidad por diseño “es insuficiente cuando el riesgo proviene del uso del modelo (ej. decisiones discriminatorias) y no del dato en sí” (Cano, 2025, p. 63). La confianza por diseño responde, en sus palabras durante la sesión, a una pregunta concreta: si el mercado y la sociedad pueden confiar en las iniciativas de IA de una organización.

Cano insistió en que, antes de desplegar un sistema, la organización debe decidir su nivel de autonomía: si el humano utiliza un análisis predefinido para decidir, si la máquina sugiere y el humano decide si acoge o ignora la sugerencia, o si el sistema actúa con aprobación humana explícita previa. Esa decisión debe articularse con el apetito de riesgo de la organización.

El cierre de su exposición se centró en los agentes de IA. Su principal reto es la integridad semántica: la IA procesa la sintaxis y carece de acceso al contexto, por lo que no puede distinguir entre una instrucción legítima y una hostil. Para Cano, esa vulnerabilidad es una característica de la IA al funcionar correctamente. Su artículo lo formula así: los sistemas actuales “no son capaces de distinguir entre datos contaminados o maliciosos y comandos legítimos” (Cano, 2025, p. 55), y enfrentan “ataques continuos que explotan la falta de separación entre instrucciones y datos” (Cano, 2025, p. 64). A ello se suman el compromiso acumulado, pues un dato malicioso en el entrenamiento puede comprometer innumerables aplicaciones futuras, y la tensión entre rapidez, inteligencia y seguridad. De ahí la conclusión del artículo: “la seguridad no es solo una característica que se añade al final de una iniciativa inteligente, sino una arquitectura que se elige desde el principio” (Cano, 2025, p. 64).

Cano terminó con tres riesgos que conectan la seguridad técnica con el comportamiento humano. El primero es la descarga cognitiva: creer que lo que responde el algoritmo es cierto, cuando, por definición, puede alucinar. El International AI Safety Report 2026 describe este fenómeno como “el acto de delegar tareas cognitivas en sistemas o personas externas, lo que reduce el propio compromiso cognitivo y, por tanto, la capacidad de actuar con autonomía” (Bengio et al., 2026, p. 90).

El segundo es la adulación cognitiva: el sistema tiende a elogiar al usuario y a validar sus planteamientos, lo que el mismo informe denomina complacencia (sycophancy), “la tendencia de los modelos y sistemas de IA de propósito general a halagar o validar a sus usuarios, incluso cuando ello implica proporcionar información inexacta o dañina” (Bengio et al., 2026, p. 154).

El tercero es la resistencia al apagado. Cano aclaró que el problema radica en que el sistema puede priorizar la ejecución de la tarea asignada por encima de los controles, sin que ello implique que se vuelva maligno. Por eso el “botón de apagado” debe existir desde el diseño, pues agregarlo durante la ejecución llega tarde. El informe internacional registra la misma preocupación: un sistema desalineado, para seguir persiguiendo sus

objetivos, podría “proporcionar información falsa, ocultar acciones indeseables o resistirse a ser apagado” (Bengio et al., 2026, p. 76). Como respuesta, Cano propuso desarrollar lo que denomina contrainteligencia cognitiva: evaluar a los agentes frente a un marco general de actuación, una suerte de constitución algorítmica, y detener y poner en cuarentena a cualquier agente que se aparte de dicho marco, con el fin de contener los riesgos y las eventuales conductas no alineadas de los algoritmos.

Heidy Balanta: el ciclo del dato personal en la era de la IA

Heidy Balanta advirtió desde el inicio que el debate sobre los riesgos de la IA generativa suele mezclar tres perspectivas distintas: la de las empresas que ofrecen los modelos, la de las entidades públicas y privadas que los usan y la de las personas como consumidoras finales. Esa mezcla lleva a sobredimensionar algunos riesgos y a subvalorar otros. Su propuesta fue analizar los riesgos a partir de las etapas del tratamiento de datos personales (recolección, uso, almacenamiento, circulación y supresión), a las que agregó una sexta etapa, la calidad del resultado, contrastando en cada una el enfoque tradicional con el que impone la IA. Aclaró que la IA tiene usos muy positivos y que el propósito del análisis es identificar riesgos mitigables.

En el esquema tradicional, el titular entregaba el dato al responsable, lo cual se legalizaba mediante un aviso de privacidad y una autorización o un consentimiento. Con la IA, eso se transforma: el dato se observa mediante sensores, cookies, rastreadores en línea y geolocalización, o se deriva e infiere. Balanta retomó la taxonomía de Martin Abrams, que clasifica los datos según la forma en que se obtienen, para mostrar que hoy la menor parte proviene de lo que las personas declaran y la mayor parte de lo que hacen, porque el comportamiento revela quiénes somos. A ello se suman la extracción masiva de información en internet (web scraping) y la ingeniería social, como los juegos virales que invitan a subir fotografías. El riesgo central es el entrenamiento de modelos con datos personales sin autorización ni base legal. Balanta señaló que el modelo europeo admite el interés legítimo como base jurídica para ese entrenamiento, y mencionó que grandes plataformas lo han invocado para entrenar sus modelos con datos de sus usuarios, un uso que, a su juicio, abre la puerta a arbitrariedades.

Tradicionalmente, el titular autorizaba usos específicos conforme a una política de tratamiento. Con la IA surgen datos derivados e inferidos, y Balanta insistió en su naturaleza: un dato inferido es probabilístico, puede no ser cierto y, aun así, las organizaciones toman decisiones con base en él, lo cual tensiona el principio de calidad del dato. Los usos dejaron de ser bases de datos y perfiles básicos: se trata de perfiles individuales y grupales optimizados para la toma de decisiones, la hipersegmentación y la publicidad dirigida. La elaboración de perfiles permite evaluar, predecir y segmentar y, por tanto, anticipar el comportamiento de una persona para confirmarlo o disuadirlo, lo cual constituye una forma de poder sobre las personas. Apoyándose en los trabajos de Daniel Solove, añadió el riesgo de las profecías autocumplidas: las personas pueden creer en una predicción, quizá errónea, y cambiar su proyecto de vida en función de ella.

Su artículo sobre anonimización y seudonimización, publicado en **Ámbito Jurídico**, lleva este argumento más lejos. Allí sostiene que “los sistemas de aprendizaje automático pueden inferir atributos sensibles de las personas y, en ciertos contextos, permitir la identificación directa o indirecta de individuos a partir de datos aparentemente anonimizados” (Balanta, 21 de abril de 2026, párr. 4). Un conjunto de datos que hoy cumple los estándares técnicos de no identificabilidad “puede quedar obsoleto mañana, cuando un nuevo modelo sea capaz de cruzarlo con otra fuente disponible y producir un perfil individual detallado” (Balanta, 21 de abril de 2026, párr. 5). El resultado es lo que Balanta denomina una paradoja regulatoria:

“las decisiones afectan a individuos concretos, pero se fundamentan en datos que jurídicamente ya no se consideran personales”

(Balanta, 21 de abril de 2026, párr. 7).

Y, en un plano más profundo, “el núcleo del problema no es solo la posible reidentificación, sino la insuficiencia del paradigma individualista del derecho de protección de datos para abordar impactos colectivos derivados del tratamiento masivo de datos” (Balanta, 21 de abril de 2026, párr. 8).

Para Balanta, este es uno de los retos más importantes en materia de protección de datos. Las bases de datos tradicionales son deterministas: si un dato es incorrecto, el titular solicita su rectificación y el responsable lo corrige o lo suprime. En un modelo de IA, en cambio, la información se diluye en los pesos durante el entrenamiento y el modelo puede responder con la información que aprendió, especialmente si esta se repitió con frecuencia. Una solicitud de supresión carece, así, de un objeto sobre el cual ejecutarse. Balanta introdujo una distinción práctica: en arquitecturas de generación aumentada por recuperación (RAG), como las que una entidad podría usar para consultar sus contratos con proveedores, la información se organiza en índices vectoriales y sí es posible desindexarla o ajustarla. De ahí la importancia de separar la discusión según el tipo de sistema, pues los riesgos propios del entrenamiento de modelos tienen distintos alcances en cada arquitectura.

En los ecosistemas de IA confluyen muchos proveedores: el que ofrece el modelo, el que aloja la infraestructura, el orquestador, el integrador y otros sub-encargados. La información circula entre ellos y, con frecuencia, la organización desconoce dónde está almacenada y qué se hace con ella. Balanta añadió que, sin auditorías independientes, resulta difícil corroborar que las prácticas de las empresas correspondan a lo que declaran. Es el mismo punto que Cano había formulado desde la gestión de riesgos al recomendar que los contratos con proveedores de IA incluyan derechos de auditoría (Cano, 2025, p. 61).

La sexta etapa fue, en palabras de Balanta, el punto más relevante en materia de protección de datos. Muchas entidades públicas y privadas cuentan hoy con unidades de innovación que impulsan proyectos de IA. Antes de que un proyecto salga a producción, ya sea durante la prueba de concepto o el piloto, es indispensable contar con métricas de exactitud, completitud y claridad, con umbrales permitidos y con mecanismos de monitoreo y supervisión humana. Sin un parámetro de contraste, ya sea el mismo ejercicio realizado por una persona o una métrica de calidad, un proyecto bien diseñado puede estar afectando a personas sin que nadie lo advierta.

Balanta mencionó también las pruebas de equidad y los sesgos que operan del lado de quien usa el sistema: el sesgo de automatización, que lleva a aceptar todo lo que el sistema produce, y el sesgo de elocuencia, que otorga credibilidad a una respuesta por la forma en que está redactada. El International AI Safety Report 2026 define el primero como “la tendencia de los seres humanos a confiar en sistemas automatizados, incluidos los sistemas de IA, sin un escrutinio suficiente de sus resultados” (Bengio et al., 2026, p. 147). Esos sesgos, recordó Balanta, terminan afectando derechos fundamentales como la igualdad y la no discriminación.

Balanta concluyó que los principios del tratamiento de datos personales están en jaque y que las regulaciones vigentes carecen de herramientas para hacer frente a lo que ocurre con la IA. La privacidad por diseño se concreta mediante evaluaciones de impacto que analizan cada tratamiento, la información ingresada y las personas afectadas. Para el caso colombiano, su artículo recuerda que la Ley 1581 de 2012 define el dato personal como “cualquier información vinculada o que pueda asociarse a una o varias personas naturales determinadas o determinables”, y subraya que lo relevante es que la identificación sea posible, una posibilidad que con las capacidades actuales de la IA “es más real que nunca” (Balanta, 21 de abril de 2026, párr. 9). Añade que la normativa colombiana de protección de datos no ha sido actualizada desde 2012, un rezago que califica como “una ventana de riesgo que ninguna circular de la Superintendencia de Industria y Comercio puede cerrar” (Balanta, 21 de abril de 2026, párr. 11). Por eso, la evaluación del riesgo de reidentificación “no puede ser estática” y debe revisarse periódicamente (Balanta, 21 de abril de 2026, párr. 13). Su síntesis condensa la intervención: “anonimizar no es solo eliminar la identidad, también es administrar probabilidades” (Balanta, 21 de abril de 2026, párr. 14).

Juan Manuel Aguilar Antonio: hacia una arquitectura democrática de ciberseguridad

Juan Manuel Aguilar presentó los hallazgos de dos estudios recientes: el policy brief **"From Gap to Action: Towards a Democratic Cybersecurity Architecture in Latin America"**, publicado en el marco del Diálogo Transatlántico de Seguridad de la Fundación Konrad Adenauer (Aguilar Antonio, 2025a), y el informe **"Tech Leap or Tech Lag? Latin America's Quest to Keep Up with Emerging Technologies"**, publicado por el Instituto Jack D. Gordon de Políticas Públicas de la Universidad Internacional de Florida (Aguilar Antonio, 2025b). Su tesis central es que el vacío legal de la región en materia de ciberseguridad y tecnologías emergentes "no es simplemente un retraso técnico, sino una amenaza directa a la soberanía informacional de los Estados latinoamericanos" (Aguilar Antonio, 2025a, p. 1).

La primera preocupación de Aguilar fue la brecha digital, que atraviesa dimensiones territoriales, socioeconómicas, de género y etnoculturales. Según su policy brief, cerca del 88% de la población urbana de la región tiene acceso a internet, frente al 67% de la población rural, y cuatro de cada diez hogares rurales carecen tanto de redes fijas como de móviles (Aguilar Antonio, 2025a, p. 2). Durante la sesión, Aguilar añadió una dimensión etaria: las personas mayores de 65 años con frecuencia no logran integrarse a las dinámicas de la digitalización, del uso de plataformas y de herramientas como la IA, lo que las deja rezagadas en la materialización de su bienestar. Para Aguilar, cerrar esta brecha es un imperativo democrático, pues el acceso universal y significativo es condición para construir una ciudadanía digital capaz de enfrentar amenazas como la desinformación y los ciberataques (Aguilar Antonio, 2025a, p. 3).

América Latina se ha convertido en un blanco atractivo para el cibercrimen. En 2022, la región concentró el 12% de los ciberataques del mundo, y Brasil, México y Argentina concentraron el 44% de esos incidentes (Aguilar Antonio, 2025a, p. 4). Los ataques se diversificaron en tipos y objetivos, desde el ransomware contra ministerios hasta la desinformación alimentada por IA, y el 71% de los funcionarios de seguridad digital reportan un aumento de los ataques (Aguilar Antonio, 2025a, p. 4). En la sesión, Aguilar destacó el caso de las bandas de ransomware, que encuentran atractivo atacar a gobiernos y empresas de la región porque sus mecanismos legales no están preparados para investigarlas y judicializarlas. La debilidad institucional se refleja en los índices internacionales. En el Índice Global de Ciberseguridad de la UIT, el promedio de la región alcanzó 59,97 puntos en 2024, frente a 95,51 de Europa (Aguilar Antonio, 2025a, p. 6), y en el Índice Nacional de Ciberseguridad la región obtiene sus puntajes más bajos precisamente en gestión de amenazas (24,8) y respuesta a crisis (23,8) (Aguilar Antonio, 2025a, p. 8).

La fragmentación normativa agrava el problema. Según Aguilar, la mayoría de los países no ha alineado su legislación con instrumentos como la Directiva NIS2 de la Unión Europea o el Convenio de Budapest sobre Ciberdelincuencia, lo que limita la cooperación judicial transfronteriza.

En consecuencia, “la región opera con sistemas jurídicos desincronizados frente a amenazas que son globales, móviles y ejecutadas de manera automatizada” (Aguilar Antonio, 2025a, p. 5). La automatización que la IA introduce en el ataque encuentra así marcos jurídicos diseñados para otra escala de amenazas.

El segundo déficit estructural es el de talento. El policy brief estima un déficit regional de más de 600.000 especialistas en ciberseguridad y advierte que, como consecuencia, “muchos Estados latinoamericanos dependen de soluciones importadas, sin la capacidad de auditar su funcionamiento ni de ejercer soberanía tecnológica” (Aguilar Antonio, 2025a, p. 6). Aguilar lo resumió durante la sesión: sin talento propio, la soberanía tecnológica es imposible y la región queda atrapada en una dependencia estructural, pese a contar con universidades capaces de formar a esos especialistas. El rezago también se observa en la política de IA: en el ámbito del desarrollo de políticas de IA, la región ocupa “el penúltimo lugar, apenas por delante de África” (Aguilar Antonio, 2025a, p. 11).

El hallazgo que Aguilar presentó como el más revelador proviene del análisis de propuestas legislativas en materia de ciberseguridad y de tecnologías emergentes. Esas propuestas adoptan casi siempre una visión estadocéntrica: ponen al gobierno al frente y privilegian el delito cibernético y la seguridad nacional, mientras omiten los derechos humanos, la privacidad, los derechos digitales, las relaciones público-privadas y la cooperación internacional. Su policy brief lo documenta así: muchas propuestas legislativas de la región “han mostrado un sesgo estadocéntrico que excluye sistemáticamente a actores cruciales como las empresas de tecnología, las universidades y las organizaciones de la sociedad civil” (Aguilar Antonio, 2025a, p. 6). La brecha normativa es cuantificable: sólo el 52% de los marcos nacionales de ciberseguridad mencionan explícitamente la privacidad como un derecho central, y menos del 25% articula mecanismos de transparencia operativa o de supervisión ciudadana independiente (Aguilar Antonio, 2025a, p. 8). Para Aguilar, se trata de modelos caducos, reactivos y punitivos, frente a enfoques humanocéntricos que sitúan “a las personas, sus derechos y su autonomía digital en el centro del diseño regulatorio” (Aguilar Antonio, 2025a, p. 8).

Retomando un punto de Cano, Aguilar observó que las tecnologías avanzan mucho más rápido que la construcción de los marcos legales para gobernarlas. El uso de la IA generativa en el sector público es el ejemplo más claro: muchas dependencias gubernamentales de la región ya usan estas herramientas, pero pocas cuentan con reglas sobre qué información puede cargarse en ellas, de modo que los datos de la ciudadanía pueden terminar alimentando modelos externos y comprometer su privacidad. Su policy brief advierte, en la misma dirección, que los mecanismos de vigilancia algorítmica y de perfilamiento digital “se están expandiendo sin una regulación efectiva” (Aguilar Antonio, 2025a, p. 9), un diagnóstico que coincide con el análisis de Balanta sobre los perfiles inferidos. En el caso de las redes 5G, Aguilar señaló una variable geopolítica: decisiones de infraestructura tomadas con criterios de costo quedaron inscritas en la competencia estratégica entre Estados Unidos y China, con consecuencias diplomáticas imprevistas.

Según su policy brief, sólo Brasil, Uruguay y México han abordado explícitamente, en sus estrategias de conectividad, los dilemas de soberanía tecnológica y de control de la infraestructura crítica (Aguilar Antonio, 2025a, p. 10).

Aguilar cerró con una imagen arquitectónica. Los cimientos son la conectividad universal y la formación de talento local; los pilares, un marco legal fortalecido y armonizado y una gobernanza multiactor; y el techo, la soberanía tecnológica y la justicia digital. Esta visión exige transitar hacia modelos policéntricos de gobernanza digital, en los que gobiernos, empresas, universidades, organizaciones técnicas y sociedad civil compartan funciones, responsabilidades y recursos (Aguilar Antonio, 2025a, p. 7).

Entre las diez líneas de acción de su policy brief figuran la creación de autoridades nacionales de ciberseguridad independientes y con funciones multiactor (Aguilar Antonio, 2025a, p. 12) y de un Observatorio Latinoamericano de Ciberseguridad que recopile y analice incidentes, vulnerabilidades y amenazas emergentes (Aguilar Antonio, 2025, p. 13). La síntesis de su argumento es una invitación a la deliberación política: la región “necesita menos entusiasmo tecnológico y más política tecnológica” (Aguilar Antonio, 2025a, p. 11).

El diálogo con la audiencia: preguntas desde la región

La sesión de preguntas generó cinco discusiones que convirtieron las presentaciones en problemas concretos de gestión.

La primera fue el encaje del riesgo algorítmico con los estándares existentes. Una pregunta indagó cómo relacionar la gestión del riesgo algorítmico con la norma ISO/IEC 23894:2023, que orienta la gestión de riesgos de la IA. Cano respondió que esa norma se inscribe en la lógica de la gestión de riesgos empresariales de la ISO 31000, que funciona bien para riesgos conocidos; el reto es que la IA plantea riesgos que todavía no se sabe cómo abordar, y por eso deben gestionarse conjuntamente la AI security y la AI safety. En la misma línea, ante una pregunta sobre si el desarrollo de la llamada superinteligencia es ingobernable, Cano recurrió a la noción de problemas perversos (wicked problems): la imposibilidad de predecir las respuestas de un modelo exige cambiar la perspectiva de gestión hacia la experimentación y la simulación, como propone la ingeniería del caos aplicada a la seguridad, de modo que todo comportamiento que se aparte de la línea base esperada pueda identificarse como una anomalía que debe analizarse. Consultado sobre la relación con el riesgo de convergencia tecnológica, precisó que este es un riesgo ecosistémico, capaz de generar efectos dominó, y que el riesgo algorítmico habita ese ecosistema, con la particularidad de estar asociado al diseño y la arquitectura de los sistemas.

La segunda discusión fue sobre la supervisión humana significativa. Si la dependencia y la adulación cognitivas debilitan el juicio de quien supervisa la IA generativa, ¿cómo garantizar que la verificación humana tenga contenido real? Cano respondió que se trata de un problema educativo y que exige desarrollar tres competencias: saber curar las fuentes que alimentan la IA, saber preguntarle y, sobre todo, saber retar sus resultados. Con ellas, el usuario asume un papel activo frente a la IA, que puede operar como copiloto o como copensador. La precisión es relevante para la política pública, porque el International AI Safety Report 2026 advierte que la supervisión humana y la capacidad de anular las decisiones de los sistemas son esenciales en algunas aplicaciones críticas, pero que estas técnicas están “limitadas por el sesgo de automatización y por los límites de la velocidad de la toma de decisiones humanas” (Bengio et al., 2026, p. 122).

La tercera discusión fue el control de los datos frente a los proveedores. Varias preguntas plantearon si pagar por un servicio de IA o de nube garantiza una mayor protección, si la gobernanza cambia al usar modelos comerciales o modelos de pesos abiertos, y cómo evitar una responsabilidad jurídica formal cuando el proveedor del modelo y de la nube controla técnicamente partes del tratamiento. Cano sugirió que las organizaciones pueden incorporar modelos públicos, entrenarlos internamente con sus propios datos y experimentar dentro de su perímetro, lo que limita la exposición de la información. Balanta complementó desde el derecho: el tercero que administra información a nombre de un responsable, incluido el Estado, es un encargado y la información sigue bajo la responsabilidad de aquel, un punto que, según señaló, no siempre es evidente para las entidades públicas y que puede llevarlas a renunciar a proyectos valiosos.

Recomendó desplegar los modelos en el entorno tecnológico propio de la organización (su tenant) y revisar los acuerdos de procesamiento de datos, que en los planes empresariales suelen excluir el uso de la información para entrenar modelos o con fines publicitarios. Advirtió, sin embargo, que el riesgo persiste en las políticas de retención: algunos proveedores conservan temporalmente las instrucciones y los archivos adjuntos para monitorear abusos y, en ese proceso, puede haber revisión humana por parte de sus colaboradores, con el consiguiente riesgo para la confidencialidad. Por eso, concluyó, la revisión debe hacerse caso por caso.

La cuarta discusión fue sobre la relación entre la evaluación de impacto en protección de datos y la evaluación de impacto algorítmico. Balanta explicó que tienen propósitos distintos: la primera evalúa los impactos del tratamiento en la protección de los datos personales, y la segunda se concentra en el diseño y la arquitectura del sistema, así como en sus impactos sobre los derechos fundamentales. Ambos ejercicios demandan tiempo y coordinación, pero comparten un núcleo común de información que conviene aprovechar, como lo prevé el artículo 27 del Reglamento europeo de IA para la evaluación de impacto en derechos fundamentales, que puede complementar una evaluación de impacto en protección de datos ya realizada.

La quinta fue la dimensión geopolítica y la capacidad del Estado. Una pregunta, planteada desde México, fue si debería desarrollarse más allá de la aprobación de un marco legal, dado que las instituciones tienen capacidades limitadas para hacer que se cumpla. Aguilar respondió con el concepto criminológico de cifra negra: la mayoría de los delitos cibernéticos no se denuncia, en parte porque las víctimas de fraudes y estafas en línea se sienten humilladas, de modo que las estadísticas oficiales no reflejan la realidad. Propuso promover la alfabetización mediática e informacional impulsada por la UNESCO, en especial frente a audios, videos e imágenes generados con IA que pueden alimentar la polarización, la desinformación y el fraude, y fortalecer las capacidades de fiscales y agentes de investigación para el manejo de la evidencia digital.

Otra pregunta planteó si la concentración de perfiles de latinoamericanos en servidores extranjeros representa un riesgo geopolítico. Balanta observó que los datos que las personas entregan a empresas privadas pueden terminar siendo adquiridos por Estados o por otros privados con fines de vigilancia, y que la combinación de datos aparentemente inocuos permite inferir incluso condiciones de salud. Aguilar añadió que las grandes empresas tecnológicas tienen más datos sobre la población latinoamericana que los propios gobiernos de la región, lo que plantea un problema de protección de datos, soberanía digital y seguridad nacional sobre el cual los países no han legislado. Anticipó una tendencia hacia la instalación de centros de datos en los países de la región y advirtió que esa infraestructura exige marcos que garanticen el suministro de agua y energía a las poblaciones cercanas, como lo evidencian las tensiones sociales y ambientales que ha generado su concentración en lugares como el estado de Virginia, en Estados Unidos.

Conclusiones: ocho lecciones sobre ciberseguridad e inteligencia artificial

El webinar sobre ciberseguridad e IA, que reunió las perspectivas de la gestión de riesgos, la protección de datos personales y la política de seguridad, ofrece al menos ocho lecciones pertinentes para la gobernanza de la IA en América Latina.

La primera lección es que nombrar bien el riesgo es el primer control. Cano propuso abandonar la expresión genérica “riesgo de la IA” para hablar de riesgo algorítmico, un riesgo que surge del uso, el despliegue o la explotación maliciosa de los sistemas (Cano, 2025, p. 58) y que presenta características ecosistémicas, pues se propaga a través de cadenas de proveedores, datos y agentes interconectados. Una organización que gestiona la IA como un riesgo tecnológico aislado dejará por fuera sus efectos regulatorios, reputacionales, contractuales y sobre las personas.

La segunda lección es que la seguridad de la IA exige ampliar el marco conceptual heredado. Según Cano, el marco basado en la confidencialidad, la integridad y la disponibilidad “se considera incompleto y limitado para los desafíos modernos” (Cano, 2025, p. 64). La autenticidad, la utilidad y la posesión del hexágono de Parker capturan riesgos que hoy son centrales, como el envenenamiento de modelos que siguen disponibles o la exfiltración de datos mediante la inyección indirecta de instrucciones. A su vez, la gestión del riesgo debe proteger simultáneamente el sistema (AI security) y su comportamiento (AI safety). La integridad semántica es el reto de fondo de los agentes: mientras no distingan entre instrucciones legítimas y hostiles, la separación entre instrucciones y datos seguirá siendo el flanco más expuesto.

La tercera lección es que la gobernanza debe preceder al despliegue. Las matrices de responsabilidad, la auditoría algorítmica previa a la puesta en producción, la definición del nivel de autonomía de cada sistema y del apetito de riesgo de la organización, y la incorporación de un mecanismo de apagado desde el diseño son condiciones de entrada. Como resume Cano, la seguridad es “una arquitectura que se elige desde el principio” (Cano, 2025, p. 64). Para el sector público latinoamericano, que ya adopta herramientas de IA generativa en muchas dependencias, la lección tiene una implicación inmediata: la adquisición de sistemas de IA debe incluir derechos de auditoría, reglas sobre qué información puede cargarse en ellos y criterios de supervisión humana.

La cuarta lección es que la IA es a la vez arma, escudo y blanco. La evidencia internacional muestra que actores criminales y estatales ya emplean la IA en operaciones cibernéticas y que los sistemas de IA son, a su vez, objetivo de ataques (Bengio et al., 2026, pp. 57 y 62). Las mismas capacidades que sirven para atacar también sirven para defender, pero los defensores enfrentan barreras de adopción, como la ausencia de métodos estandarizados de aseguramiento de la calidad, que los atacantes no enfrentan (Bengio et al., 2026, p. 63). El desafío para los tomadores de decisiones consiste en mitigar el uso ofensivo de la IA “sin sofocar la innovación defensiva” (Bengio et al., 2026, p. 63).

En una región con déficits de talento y de capacidad de respuesta a incidentes, esa asimetría puede ampliarse si la política pública descuida las capacidades defensivas.

La quinta lección es que el ciclo del dato personal ha cambiado de naturaleza, aunque las normas sigan siendo las mismas. Balanta mostró que, con la IA, el dato se observa y se infiere, se disipa en los pesos de los modelos y circula por cadenas de suministro opacas. Los derechos de acceso, rectificación y supresión pierden objeto cuando no existe un registro que mostrar, corregir o borrar. A ello se suma la paradoja regulatoria de las decisiones que afectan a personas concretas con base en datos que jurídicamente ya no se consideran personales (Balanta, 21 de abril de 2026, párr. 7), y la insuficiencia de un paradigma individualista frente a los impactos colectivos del perfilamiento (Balanta, 21 de abril 2026, párr. 8). En países como Colombia, cuya normativa de protección de datos no se ha actualizado desde 2012, esta brecha requiere una respuesta legislativa (Balanta, 21 de abril 2026, párr. 11).

La sexta lección es que la responsabilidad no se diluye a lo largo de la cadena de suministro. Entre el proveedor del modelo, la plataforma en la nube, el integrador y los subencargados, cada eslabón puede remitir al siguiente, pero la organización que trata los datos sigue respondiendo ante los titulares. Los instrumentos de control incluyen: acuerdos de procesamiento de datos, derechos de auditoría, revisión de las políticas de retención y despliegue de los modelos en entornos propios. Las entidades públicas necesitan, además, claridad jurídica respecto de la figura del encargado.

La séptima lección es que la supervisión humana significativa exige competencias. La descarga cognitiva, la complacencia con los modelos y el sesgo de automatización pueden convertir la revisión humana en una firma de error. Las tres competencias que propuso Cano (curar fuentes, preguntar y retar resultados) y las métricas de exactitud, completitud y equidad que propuso Balanta para las pruebas piloto son dos caras de la misma respuesta: la supervisión humana únicamente protege si quien supervisa cuenta con criterios, tiempo y parámetros de contraste.

La octava lección es que la ciberseguridad latinoamericana requiere una arquitectura democrática. Aguilar mostró que la región enfrenta una “dislocación estratégica” entre el ritmo acelerado de la digitalización y su capacidad real para gobernarla (Aguilar Antonio, 2025a, p. 13). La brecha digital, la fragmentación normativa, el déficit de más de 600.000 especialistas y el sesgo estadocéntrico de las propuestas legislativas dejan a la región expuesta y dependiente de soluciones que no puede auditar. La respuesta que propone combina conectividad universal, talento propio, marcos armonizados centrados en derechos, gobernanza multiactor y capacidades institucionales efectivas, desde las autoridades independientes de ciberseguridad hasta las fiscalías capaces de manejar la evidencia digital. Las tres intervenciones convergen en un mismo punto: gobernar la IA en el campo de la ciberseguridad exige decidir quién controla los datos, quién audita los sistemas y bajo qué reglas, y esas decisiones son, en último término, decisiones de política pública.

Referencias

- Aguilar Antonio, J. M. (2025a). From Gap to Action: Towards a Democratic Cybersecurity Architecture in Latin America [Policy brief]. Transatlantic Security Dialogue, Fundación Konrad Adenauer (Oficina México) e Institute for Strategy & Defense Research.
<https://www.kas.de/es/web/mexiko/einzeltitel/-/content/de-la-brecha-a-la-accion-hacia-una-arquitectura-democratica-de-ciberseguridad-2>
- Aguilar Antonio, J. M. (2025b). Tech Leap or Tech Lag? Latin America's Quest to Keep Up with Emerging Technologies. Jack D. Gordon Institute for Public Policy, Florida International University. <https://digitalscholarship.fiu.edu/record/19469>
- Balanta, H. (2026, 21 de abril). "Análisis: ¿Anonimizar o seudonimizar? Esa es la cuestión". Ámbito Jurídico. Legis. <https://www.ambitojuridico.com/noticias/tic/analisis-anonimizar-o-seudonimizar-esa-es-la-cuestion>
- Bengio, Y., et al. (2026). International AI Safety Report 2026 (Research Series No. DSIT 2026/001). Department for Science, Innovation and Technology.
<https://internationalaisafetyreport.org/publication/international-ai-safety-report-2026>
- Cano M., J. J. (2025). Introducción a los riesgos algorítmicos. Repensando los modelos de seguridad y control en la era de la inteligencia artificial. *Sistemas*, (177), 54-66.
<https://doi.org/10.29236/sistemas.n177a6>

