Gobernanza de la inteligencia artificial generativa

Una introducción a la confianza por diseño.

DOI: 10.29236/sistemas.n179a7

Resumen

Este artículo propone el marco de "Confianza por Diseño" (CpD) como estrategia de gobernanza para la IA Generativa (IA Gen). Ante los vacíos de los enfoques tradicionales, la CpD integra la seguridad, la privacidad y la ética. El modelo busca armonizar los intereses del Estado (garante), las empresas (innovación) y los individuos (autonomía cognitiva). Se enfatiza que la sostenibilidad tecnológica depende de salvaguardas proactivas que aseguren la integridad algorítmica y mitiguen riesgos como la "*hipersuación*"[1] y la "infodemia". La reflexión concluye que la CpD permite una innovación responsable, transformando la IA en un habilitador de bienestar social y rentabilidad económica mediante una colaboración pluralista entre los diferentes actores del ecosistema digital soporte de la IA.

Palabras clave

Confianza por diseño, seguridad por diseño, privacidad por diseño, ética, IA Generativa

1 La "hipersuasión" (en inglés: *hypersuasion*) es un concepto acuñado por Luciano Floridi para describir formas de persuasión altamente sofisticadas y personalizadas habilitadas por sistemas de IA. Fuente: Floridi, L. (2024). Hypersuasion – On AI's persuasive power and how to deal with it. *Philosophy & Technology*, *37*(64). https://doi.org/10.1007/s13347-024-00756-6

Introducción

La inteligencia artificial (IA) y el aprendizaje automático han transformado radicalmente el panorama de los negocios y la dinámica del mundo actual. La tasa de adopción y penetración ha venido en un crecimiento sostenido desde el lanzamiento de ChatGPT en 2022, cuando alcanzó los 100 millones de usuarios activos mensuales en solo dos meses, convirtiéndose en la aplicación de consumo de más rápido crecimiento en la historia hasta ese momento (Floridi, 2023).

De otra parte, se registró un aumento masivo del 425% en las inversiones de capital de riesgo destinadas a la IA Gen desde el año 2020 hasta mediados de 2023. De igual forma, se estima que el mercado de la IA Gen se convertirá en una industria de 1,3 billones de dólares para el año 2032, lo que sugiere que las organizaciones continuarán desarrollando iniciativas para incorporar esta tecnología procurando mayor automatización y transformación de los diferentes sectores productivos a nivel internacional (Habuka & Socol de la Osa, 2025).

Esta carrera tecnológica necesariamente genera tensiones entre diferentes intereses en la dinámica de la sociedad actual mediada por tecnologías disruptivas y emergentes, creando vacíos de responsabilidad y riesgos emergentes, que no se pueden explicar o atender exclusivamente desde las perspectivas tradicionales como la seguridad y la privacidad, las cuales están diseñadas para contextos específicos de los retos y exigencias, tanto de los reguladores como de los individuos, asociados con estándares y buenas prácticas reconocidas.

Por tanto, este documento propone un modelo de “confianza por diseño” como una estrategia para la gobernanza corporativa de la IA Generativa (IA Gen). A través de un análisis de los intereses divergentes y convergentes de las empresas, los reguladores y los ciudadanos, se argumenta que la sostenibilidad a largo plazo de la IA depende de la integración proactiva de controles éticos y técnicos desde la fase de diseño, con el fin de limitar los daños por fallas propias del funcionamiento del ecosistema digital donde opera y así, asegurar su aceptación social.

Para ello, este trabajo inicia con una breve revisión de literatura de los conceptos de privacidad por diseño y seguridad por diseño como referentes básicos disponibles en la actualidad. Luego se plan-

tea la paradoja de la IA Gen en el contexto actual, a reglón seguido se detallan las tensiones entre los tres participantes claves de la dinámica social alrededor de la IA Gen como son el mercado, el estado y los individuos para establecer las bases de la construcción del modelo de confianza por diseño. Seguidamente se detalla la propuesta de la confianza por diseño y se finaliza con algunas conclusiones y retos para este modelo ahora y en el futuro.

La paradoja de la IA Gen: más humanidad que tecnología

La paradoja de la IA Gen establece que, a medida que la tecnología de procesamiento de lenguaje se vuelve más sofisticada y capaz de operar a escala, el éxito de su implementación depende más críticamente de factores humanos que la máquina no puede resolver: la capacidad de comprender, cuestionar e integrar sus resultados (Sieber, 2026).

Aunque la IA Gen se clasifica como una IA estrecha, diseñada para tareas específicas mediante algoritmos estadísticos (Abaimov & Martellini, 2022), su adopción masiva ha generado una zona de la desilusión productiva, donde las empresas invierten grandes sumas sin transformar realmente sus modelos de negocio.

Desde la perspectiva del riesgo cibernético, la paradoja se agrava por el fenómeno de la "adopción en la sombra" (*Shadow AI*), donde los empleados utilizan herramientas personales, exponiendo datos corporativos y comprometiendo el cumplimiento regulatorio (Sieber, 2026). Esto es particularmente peligroso debido a la naturaleza de "caja negra" de los modelos de aprendizaje profundo (*Deep Learning*), cuya opacidad dificulta anticipar decisiones o detectar manipulaciones externas como el envenenamiento de datos (*data poisoning*) (Abaimov & Martellini, 2022).

Por tanto, la transición de "operadores" a "orquestadores" requiere una humildad tecnológica que reconozca que la verificación humana debe ser una parte estructural del flujo de trabajo, no un control final. Mientras el operador tradicional gestiona herramientas de forma rutinaria, el orquestador diseña la interacción estratégica con sistemas autónomos. Esto es, se delega en la máquina la capacidad analítica y ejecución a escala, reservando para el juicio humano la ética, la empatía y la gestión de la ambigüedad. Así, el valor real surge de capturar el "dividendo de la aumentación" (valor creado de maneras que ni el ser humano ni las máquinas podrían lograr en solitario) mediante una supervisión humana estructural (Sieber, 2026).

En la Tabla 1 se presenta un cuadro resumen de los retos de la paradoja de la IA Gen.

Tabla 1. *Retos de la paradoja de la IA Gen*

Fiabilidad Técnica	Las "alucinaciones" son fallos estructurales; la IA Gen genera información plausible pero falsa, requiriendo validación constante (Sieber, 2026).
Ciberseguridad	La IA es de doble uso: las mismas herramientas que defienden pueden usarse para ataques adversarios y generación de malware elusivo (Abaimov & Martellini, 2022).
Psicológico	Resistencia al cambio ante el miedo a la obsolescencia, la pérdida de control y el fracaso público (Sieber, 2026).
Gobernanza	Necesidad de pasar de un enfoque basado solo en eficiencia a uno de "dividendo de aumentación", donde la IA amplifica el juicio ético humano.
Explicabilidad	Superar la falta de transparencia de los modelos no lineales para construir sistemas confiables y auditables (Abaimov & Martellini, 2022)

Nota: Elaboración propia basado en Sieber, 2026 y Abaimov & Martellini, 2022.

Seguridad y privacidad por diseño. Un resumen de las posturas tradicionales

Los pilares teóricos de la Seguridad por Diseño (SpD) se remontan a 1975, cuando Jerome Saltzer y Michael Schroeder (1975) publicaron los principios de diseño que aún hoy constituyen el fundamento de los sistemas confiables. Su visión subrayaba que la robustez de un sistema no debe depender del secreto de su arquitectura (seguridad por oscuridad, no es seguridad), sino de la solidez de sus mecanismos intrínsecos de protección. A continuación se detalla en la tabla No.2 el resumen de las consideraciones de Saltzer y Schroeder (1975).

La implementación efectiva de la SpD en entornos modernos implica desplazar las actividades de seguridad hacia las etapas más tempranas de planificación y diseño de las soluciones informáticas. Esto incluye el modelado de amenazas proactivo, el análisis estático y dinámico de código, y la adopción de metodologías DevSecOps para automatizar la verificación de seguridad en flujos de trabajo ágiles. Desde una perspectiva económica, la literatura indica que resolver fallas de seguridad en la etapa de diseño puede reducir los costos de remediación post-producción, al tiempo que fortalece la confianza de los interesados y la resiliencia operativa frente a amenazas persistentes (NIST, 2022; Valdés-Rodríguez et al., 2023).

De otra parte, la Privacidad por Diseño (PpD) es un marco concep-

Tabla 2. *Principios de diseño seguro*

Principio	Definición	Impacto
Mínimo privilegio	Restricción de acceso a los permisos básicos necesarios para una función.	Minimiza el radio de impacto ante un posible compromiso de cuenta.
Mediación completa	Verificación sistemática de cada intento de acceso a cada objeto.	Previene la existencia de rutas de acceso no supervisadas.
Diseño abierto	La seguridad no debe depender del desconocimiento del atacante sobre el diseño.	Facilita la auditoría externa y la identificación colectiva de fallos.
Economía del mecanismo	Mantener el diseño del sistema de protección lo más simple y pequeño posible.	Reduce la probabilidad de errores de implementación y facilita la verificación.
Valores predeterminados seguros	Denegación de acceso por defecto; el permiso debe ser una excepción explícita.	Asegura que un error de configuración no resulte en una apertura no deseada.
Separación de privilegios	Requiere más de una condición o llave para acceder a recursos críticos.	Evita que la vulnerabilidad de un solo control comprometa todo el sistema.
Aceptabilidad psicológica	El diseño debe ser intuitivo para que el usuario no eluda los controles.	Fomenta el cumplimiento natural de las normas de seguridad.

Nota: Elaboración propia basada en Saltzer & Schroeder, 1975

tual y metodológico que propone la integración de la protección de la privacidad como un requisito funcional y estructural desde la fase de concepción de cualquier sistema, tecnología o práctica organizacional. Desarrollado originalmente por la Dra. Ann Cavoukian en la década de 1990, el PpD traslada la responsabilidad de la privacidad de un modelo basado exclusivamente en el cumplimiento legal reactivo a un paradigma de ingeniería proactivo.

Este enfoque se articula a través de siete principios fundamentales que rigen el ciclo de vida del desarrollo: (Cavoukian, 2009)

- *Proactivo, no reactivo*: Anticipa riesgos antes de que ocurran fallos de privacidad.
- *Privacidad como configuración predeterminada*: Los sistemas deben proteger los datos automáticamente, sin requerir acción del usuario (Privacy by Default).
- *Privacidad integrada en el diseño*: La protección no es un "añadido", sino un componente esencial de la arquitectura.
- *Funcionalidad total* (Suma positiva): Rechaza falsas dicotomías

entre seguridad y funcionalidad, buscando soluciones donde ambos objetivos coexistan.
- *Seguridad de extremo a extremo*: Asegura la protección de los datos desde su recolección hasta su eliminación segura.
- *Visibilidad y transparencia*: Las operaciones deben ser verificables y abiertas al escrutinio independiente.
- *Respeto por la privacidad del usuario*: Centra el diseño en las necesidades y el empoderamiento del individuo.

En el contexto contemporáneo, la PpD ha trascendido el ámbito teórico para convertirse en un estándar legal global, siendo el pilar central del Artículo 25 del Reglamento General de Protección de Datos (GDPR) de la Unión Europea bajo el término "Protección de datos desde el diseño y por defecto". La literatura reciente destaca que el éxito del PpD en entornos de inteligencia artificial y desarrollo ágil depende de la identificación de vulnerabilidades de privacidad en las etapas más tempranas de los requisitos de software, reduciendo costos de reparación post-producción hasta en un 90% (Del-Real et al., 2025).

La investigación técnica y jurídica demuestra de forma particular que la seguridad y la privacidad por diseño no son obstáculos a la innovación, sino cimientos fundamentales para las tecnologías inteligentes. En un mundo donde la IA Gen media en casi toda interacción informativa, la responsabilidad de los arquitectos de sistemas es equiparable a la de los legisladores. Por tanto, la defensa del futuro digital descansa sobre tres pilares: (Paseri & Durante, 2025)
- Preservación del capital semántico: La defensa de la capacidad humana para generar significado frente a la imitación sintáctica.
- Salvaguarda de la autonomía humana: asegurar la soberanía cognitiva frente a la manipulación computacional que perfila y persuade para cambiar comportamientos.
- Equidad sistémica: Asegurar que los beneficios de la IA Gen no intensifiquen la asimetría de poder o la exclusión social.

La dependencia tecnológica excesiva, en ausencia de estos tres pilares, conlleva una subordinación a poderes técnicos opacos que pueden desestabilizar la soberanía estatal y la autonomía individual. Como advierte Zou et al. (2025), el diseño es hoy una forma de legislación por código.

Iniciativas de inteligencia artificial: tensiones entre el Estado, el mercado y el individuo

La irrupción de la Inteligencia Artificial Generativa (IA Gen) en el tejido socio-técnico contemporáneo no representa sólo una optimización de la capacidad de cómputo o un refinamiento incremental de los modelos de aprendizaje automáti-

co. Constituye, en esencia, un desafío existencial[2] para los marcos legales y prácticos tradicionales de seguridad y privacidad, forzando una reevaluación de la soberanía cognitiva y la autonomía individual.

En este sentido, los intereses de las empresas por la incorporación de mayor innovación y rentabilidad, establecen nuevas propuestas y retos en el uso de la IA Gen que llevan a las organizaciones a tomar mayores riesgos para los cuales generalmente no se encuentran preparadas, obligando a las áreas jurídicas a buscar estrategias que permitan minimizar los riesgos legales y reputacionales, en una perspectiva de cumplimiento basado en el pasado: luego de que ya pasaron los hechos y buscar asumir la menor responsabilidad posible (Coles-Kemp & Burdon, 2025).

Por otro lado, está el Estado en su función de garante de los derechos individuales, de la seguridad nacional, el orden público y el cumplimiento de la ley. En este contexto, busca establecer marcos de trabajo lo suficientemente amplios y exigentes para limitar riesgos sistémicos que puedan afectar la dinámica de las infraestructuras críticas y la gobernabilidad de un país, así como evitar abusos de poder por parte de las grandes empresas de tecnología, procurando un uso limitado y vigilado de las tecnologías que permita a los individuos sentirse tranquilos por las implementaciones de iniciativas basadas en IA Gen (Coles-Kemp & Burdon, 2025; Taddeo et al., 2019).

Finalmente y no menos importante, los individuos como parte esencial de la sociedad, demandan transparencia sobre cómo se usan sus datos y si están interactuando con una máquina o un humano, con el fin de limitar los posibles efectos de una manipulación dirigida, o el uso no autorizado de sus datos personales para alimentar un modelo de IA Gen que pueda, no sólo comprometer su privacidad o buen nombre, sino crear escenarios de suplantación, o ser utilizados para distorsionar el comportamiento de un grupo de personas con fines contrarios a la Constitución y la ley (Coles-Kemp & Burdon, 2025).

Por tanto, la gobernanza de la IA Gen exige un cambio de paradigma hacia una responsabilidad digital colectiva e híbrida, que distribuya las obligaciones entre los creadores de los modelos y quienes los despliegan en el mercado. Solo a través de esta convergencia se podrá atender la “brecha de responsabilidad” que surge cuando los sistemas autónomos actúan sin una supervisión humana significativa (Coles-Kemp & Burdon, 2025). Al integrar controles éticos y técnicos desde el inicio, las organizaciones no solo aseguran el cumpli-

2 En el ámbito de la IA, esto implica la creación de una superinteligencia que supere el control humano o el despliegue de armas autónomas letales amenazando la existencia de la humanidad (Zou et al., 2025)

miento normativo, sino que construyen el capital semántico de confianza (intervención humana), que da sentido y contexto, necesario para que la IA Gen sea aceptada como una herramienta de bienestar común y no como un riesgo existencial para la humanidad.

Confianza por diseño: hacia un marco de trabajo de intereses convergentes.

La intersección donde convergen los tres intereses previamente comentados es la Confianza por Diseño (CpD). En este punto, la tecnología deja de ser un factor de incertidumbre para convertirse en un habilitador de ventajas competitivas: es lícita al cumplir con la regulación estatal, ética al respetar la autonomía del individuo y estratégica al generar valor sostenible para la empresa. Por tanto, la CpD se configura como una práctica para integrar gobernanza, ética y rendición de cuentas en la base misma de las tecnologías, asegurando que el comportamiento confiable sea el valor predeterminado y no una excepción.

Las empresas deben transitar de una innovación "sin límites" a una "responsable". La gobernanza de la IA ahora requiere directivos competentes en tecnología que puedan supervisar la "integridad de la IA" como parte de sus deberes fiduciarios (Coles-Kemp & Burdon, 2025). Cuerpos de gobierno que reconozcan la inevitabilidad de la falla de las iniciativas basadas en inteligencia artificial, la opacidad algorítmica, y por lo tanto, exijan la mediación humana en las respuestas que puedan afectar negativamente a los diferentes grupos de interés, así como protocolos de actuación definidos y practicados, cuando la IA Gen genere acciones adversas por fallas estructurales que perjudiquen directamente a sus clientes (Abaimov & Martellini, 2022).

De otro lado, las regulaciones como el Reglamento de IA de la Unión Europea clasifican los sistemas según su impacto en los derechos fundamentales. Clasifica los sistemas en cuatro categorías: inaceptable, alto, limitado y mínimo. Los de riesgo *inaceptable* están prohibidos, abarcando la afectación social y la manipulación conductual. Los sistemas de *alto* riesgo, aplicados en infraestructuras críticas, salud, educación y justicia, requieren una evaluación rigurosa, transparencia y supervisión humana. El riesgo *limitado* exige transparencia (como en chatbots), mientras que el riesgo *mínimo* (filtros de spam) no se regula. El Estado actúa como garante para evitar riesgos sistémicos y la manipulación de la opinión pública, que limite la erosión de la autonomía individual y el Estado de derecho (EUPC, 2024).

Y finalmente el individuo, como "consumidor algorítmico", puede estar perdiendo el control por la emergencia del fenómeno de la "hipersuasión", la capacidad de la

IA para manipular el comportamiento humano de forma sutil y personalizada, que amenaza la autodeterminación mental de los ciudadanos, lo que exige un diseño técnico que respete la dignidad humana y los derechos fundamentales. Un ejercicio de cuidado y responsabilidad que las organizaciones deben atender de forma proactiva para asegurar una implementación confiable y socialmente aceptada por los diferentes grupos de interés (Poncibò, 2025).

Para lograr una ventaja competitiva sostenible con IA, las organizaciones deben integrar capacidades operativas únicas. Según Waltzman et al. (2020), el éxito no reside solo en la investigación, sino en la transición efectiva de la IA a aplicaciones específicas y en procesos robustos de verificación y validación. Abaimov y Martellini (2022) destacan que competitividad de la IA surge de procesar grandes volúmenes de datos para personalizar servicios y ganar eficiencia operativa. Además, priorizar la calidad del dato permite a empresas pequeñas reducir costos frente a grandes plataformas (Papadopoulos, 2025). Finalmente, la sostenibilidad de esta ventaja requiere una gobernanza que asegure la responsabilidad digital y la seguridad, transformando la confianza en un activo estratégico (Coles-Kemp & Burdon, 2025).

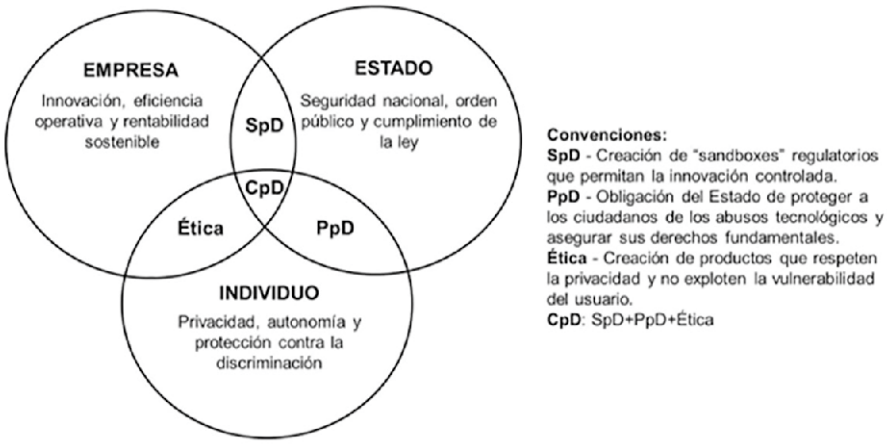
En resumen podríamos definir la confianza por diseño como un constructo relacional sociotécnico que armoniza las dimensiones de seguridad, ética y legalidad en el ecosistema de la IA, un ejercicio de equilibrio dinámico donde el mercado innova, el Estado protege y la sociedad participa, disminuyendo las lagunas de responsabilidad mediante una colaboración pluralista que trasciende la sola eficiencia técnica. En términos prácticos la CpD es la suma de la SpD+PpD+Ética, donde los requisitos de confiabilidad, explicabilidad, equidad, privacidad, seguridad y ética deben ser consideraciones de primer orden en el diseño arquitectónico de los sistemas inteligentes. La figura 1 detalla y resume el marco de trabajo detallado en esta sección.

Conclusiones

Hacia el futuro, la gobernanza de la IA Gen enfrenta desafíos claves. El primero es cerrar la “brecha de responsabilidad”, donde la complejidad de la cadena de suministro dificulta la atribución de daños causados por sistemas autónomos (Coles-Kemp & Burdon, 2025). A medida que los modelos desarrollan comportamientos emergentes impredecibles, los marcos legales tradicionales de negligencia y causalidad se ven tensionados, exigiendo nuevas formas de responsabilidad híbrida entre humanos y máquinas.

Un segundo reto crítico es la estandarización global. El reto de una

Figura 1. *Confianza por diseño*



Nota: Elaboración propia

visión homogénea de la IA enfrenta obstáculos críticos por la fragmentación de criterios nacionales e internacionales, generando un "mosaico global" de regulaciones que entorpece la innovación. El acelerado avance tecnológico supera los tiempos legislativos, impidiendo alcanzar la madurez necesaria para consensuar normas técnicas universales (Del-Real et al., 2025). Además, la opacidad algorítmica y la variabilidad del aprendizaje automático dificultan establecer métricas de seguridad y fiabilidad que sean aceptadas internacionalmente.

El tercer reto relevante es el riesgo de una "infodemia", ese flujo abundante de información que satura el ecosistema digital con desinformación automatizada y contenidos sintéticos, que impide distinguir hechos reales de datos fabricados, erosionando la confianza social en la tecnología y el conocimiento científico, exige mecanismos robustos de autenticación de contenido y marcas de agua digitales, los cuales deben asumirse como nuevas prácticas y estándares que enfrenten el reto de la integridad de la información, para configurar un estándar de debido cuidado de las empresas frente al despliegue de aplicaciones basadas en inteligencia artificial generativa (Abaimov & Martellini, 2022).

De esta forma, la Confianza por Diseño se debe configurar como un constructo jurídico de construcción colectiva que armoniza los intereses del mercado, el Estado y los individuos. Esta convergencia exige transitar hacia marcos constitutivos basados en la colaboración y el consenso entre las partes (Coles-Kemp & Burdon, 2025). Para

las empresas, implica incluir la integridad algorítmica en sus deberes fiduciarios, alineándose con la protección de derechos que el Estado asegura mediante la reglamentación vigente para la IA, y la apropiación de la ciudadanía, basada en la transparencia y la preservación de la autonomía frente a la "*hipersuación*".

Finalmente, la confianza por diseño es una propuesta para que la IA sea socialmente aceptable, tecnológicamente confiable y económicamente rentable. Solo mediante la incorporación proactiva de prácticas éticas y defensas técnicas desde la fase de diseño, junto con una conversación tripartita entre el Estado, el mercado y los ciudadanos, las organizaciones podrán navegar un futuro marcado por la incertidumbre, transformando los riesgos emergentes de la IA Gen, en un habilitador de progreso humano y bienestar para todos.

Referencias

Abaimov, S., & Martellini, M. (2022). *Machine learning for cyber agents: Attack and defence*. Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-030-91585-8

Cavoukian, A. (2009). Privacy by Design: The 7 Foundational Principles. Information and Privacy Commissioner of Ontario. https://www.ipc.on.ca/sites/default/files/legacy/2018/01/pbd-1.pdf

Coles-Kemp, L., & Burdon, M. (2025). *Understanding digital responsibilities*. Bristol University Press. https://doi.org/10.51952/9781529249798

Del-Real, C., De Busser, E., & van den Berg, B. (2025). A systematic literature review of security and privacy by design principles, norms, and strategies for digital technologies. *International Review of Law Computers & Technology*, *39*(3), 374–405. https://doi.org/10.1080/13600869.2025.2457227

European Parliament & Council of the European Union - EUPC (2024). *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. Official Journal of the European Union. http://data.europa.eu/eli/reg/2024/1689/oj

Floridi, L. (2023). AI as agency without intelligence: On ChatGPT, large language models, and other generative models. *Philosophy & Technology*, *36*(1), 1-15. https://doi.org/10.1007/s13347-023-00621-y

Habuca, H. & Socol de la Osa, D. (2025). Shaping Global AI Governance. A Path for the G7 to Foster Rule of Law in a World of Uncertainty. En Zou, M., Poncibò, C.,Ebers, M. & Calo, R. (Eds.), *The Cambridge Handbook of Generative AI and the Law*. Cambridge University Press.

National Institute of Standards and Technology (NIST). (2022). Engineering Trustworthy Secure Systems (NIST Special Publication 800-160, Vol. 1, Rev. 1). *U.S. Department of Commerce*. https://doi.org/10.6028/NIST.SP.800-160v1r1

Papadopoulos, S. (2025). Redefining rivalry: Generative AI and the evolving

landscape of competition law. En M. Zou, C. Poncibò, M. Ebers, & R. Calo (Eds.), *The Cambridge Handbook of Generative AI and the Law*. Cambridge University Press.

Paseri, L., & Durante, M. (2025). Normative and ethical dimensions of generative AI: From epistemological considerations to societal implications. En M. Zou, C. Poncibò, M. Ebers, & R. Calo (Eds.), *The Cambridge Handbook of Generative AI and the Law*. Cambridge University Press.

Poncibò, C. (2025). Regulating hypersuasion. En M. Zou, C. Poncibò, M. Ebers, & R. Calo (Eds.), *The Cambridge Handbook of Generative AI and the Law*. Cambridge University Press.

Saltzer, J. H., & Schroeder, M. D. (1975). The protection of information in computer systems. *Proceedings of the IEEE. Institute of Electrical and Electronics Engineers*, *63*(9), 1278–1308. https://doi.org/10.1109/proc.1975.9939

Sieber, S. (2026). La paradoja de la IA generativa: por qué más tecnología requiere más humanidad. *Harvard Deusto*, (358), 26-41. https://www.harvard-deusto.com/la-paradoja-de-la-ia-generativa-por-que-mas-tecnologia-requiere-mas-humanidad

Taddeo, M., McCutcheon, T., & Floridi, L. (2019). Trusting artificial intelligence in cybersecurity is a double-edged sword. *Nature Machine Intelligence*, *1*(12), 557–560. https://doi.org/10.1038/s42256-019-0109-1

Valdés-Rodríguez, Y., Hochstetter-Diez, J., Díaz-Arancibia, J., & Cadena-Martínez, R. (2023). Towards the integration of security practices in agile software development: A systematic mapping review. *Applied Sciences*, *13*(7), 4578. https://doi.org/10.3390/app13074578

Waltzman, R., Ablon, L., Curriden, C., Hartnett, G. S., Holliday, M. A., Ma, L., Nichiporuk, B., Scobell, A., & Tarraf, D. C. (2020). Maintaining the competitive advantage in artificial intelligence and machine learning. *RAND Corporation*. https://www.rand.org/pubs/research_reports/RRA200-1.html

Zou, M., Poncibò, C., Ebers, M., & Calo, R. (Eds.). (2025). *The Cambridge handbook of generative AI and the law*. Cambridge University Press. https://doi.org/10.1017/9781009492553

Jeimy J. Cano M., Ph.D, CFE, CICA. *Ingeniero y Magíster en Ingeniería de Sistemas y Computación, Universidad de los Andes. Especialista en Derecho Disciplinario, Universidad Externado de Colombia; Ph.D en Business Administration, Newport University, CA. USA. y Ph.D en Educación, Universidad Santo Tomás. Profesional certificado como Certified Fraud Examiner (CFE), por la Association of Certified Fraud Examiners y Certified Internal Control Auditor (CICA) por The Institute of Internal Controls. Profesor Distinguido de la Facultad de Derecho, Universidad de los Andes. Es director de la Revista SISTEMAS de la Asociación Colombiana de Ingenieros de Sistemas –ACIS–.*