No. 177 Octubre - Diciembre DOI: 10.29236/sistemas ISSN 0120-5919

SISTEMAS

AI Generativa: Conceptos, Retos y Promesas

ACIS
ASOCIACIÓN COLOMBIANA DE INFORMÁTICA, SISTEMAS Y TECNOLOGÍAS AFINES
Calle 93 No. 13 - 32 of. 102
Bogotá, D.C.
www.acis.org.co

ACIS

ASOCIACIÓN COLOMBIANA DE INFORMÁTICA, SISTEMAS Y TECNOLOGÍAS AFINES

ACIS Conecta

*Desde **ACIS** queremos invitarte a ser parte de ACIS Conecta, una nueva iniciativa creada especialmente para nuestros asociados.*

Actualiza tu perfil, destaca tus competencias y prepárate para conectarte con nuevas oportunidades profesionales. Gracias a nuestras alianzas con el Gobierno, el sector empresarial y organismos internacionales, podrás acceder a proyectos, vacantes especializadas y programas de formación que impulsarán tu crecimiento.

¡Aplica a nuestra encuesta!

acis.org.co/acis-conecta

301 5530540

administrador@acis.org.co

www.acis.org.co

En esta edición

Publicación de la Asociación Colombiana de Informática, Sistemas y Tecnologías Afines (ACIS)
Resolución No. 003983 del Ministerio de Gobierno
Tarifa Postal Reducida Servicios Postales Nacional S.A. No. 2015-186 4-72
ISSN 0120-5919
Apartado Aéreo No. 94334
Bogotá D.C., Colombia

Octubre - Diciembre 2025
Calle 93 No.13 - 32 Of. 102
Teléfonos 616 1407 - 616 1409
A.A. 94334
Bogotá D.C.
www.acis.org.co

Inteligencia artificial generativa

DOI: 10.29236/sistemas.n177a1

Amenazas y oportunidades en un mundo NAVI

Jeimy J. Cano M.

El contexto de un mundo NAVI (No lineal, Acelerado, Volátil e Interconectado) (McCowan et al., 2025) no anuncia una época de cambios, sino un cambio de época. Un momento en la historia de la humanidad y un punto de inflexión donde la inteligencia artificial (IA) cambia drásticamente la sociedad de forma global, transformando cada sector e industria retando su conocimiento y abriendo nuevas oportunidad de eficiencia y renovación.

Dentro de este panorama, los Modelos Grandes de Lenguaje (LLMs, (*Large Language Models*) por sus siglas en inglés), como los popularizados por ChatGPT de OpenAI, Gemini de Google o Claude de Anthropic son una manifestación específica y concreta de la Inteligencia Artificial Generativa (IA Gen). Los LLMs son modelos de redes neuronales profundas basados en la arquitectura del *transformer*. Estos modelos marcan una nueva era para el Procesamiento del Lenguaje Natural (NLP, *Natural Language Processing*), superando a los métodos tradicionales que antes solo eran diseñados para tareas de ca-

tegorización simple, pero que no podían manejar el entendimiento complejo y la generación coherente de texto (Alammar & Grootendorst, 2024).

Los LLMs logran esta capacidad avanzada al ser entrenados con extensas cantidades de datos de texto, en muchas ocasiones con grandes porciones de todo el texto disponible públicamente en Internet.

Esto les permite capturar información contextual profunda y sutilezas del lenguaje humano. Es fundamental aclarar que, si bien estos modelos tienen capacidades notables para comprender, generar e interpretar el lenguaje humano, cuando se dice que "entienden", se refiere a que procesan y generan texto de manera coherente y contextualmente relevante, no a que poseen conciencia humana.

A continuación se presenta un resumen de algunas de sus ventajas y limitaciones actuales que enmarcan las reflexiones que se plantean y desarrollan en el desarrollo de este número. (Tabla 1)

Tabla 1. Algunas ventajas y limitaciones actuales de la IA GEN

Categoría	Ventajas	Limitaciones
Eficiencia y productividad	Transformación de la Velocidad: La IA ayuda a realizar la mayoría de las tareas de forma más rápida, económica y eficiente.	Alucinación: Los LLMs pueden generar información incorrecta o contenido poco confiable.
Innovación y ventaja competitiva	Toma de decisiones estratégicas: Reunir datos complejos y sugerir cursos de acción clave para las empresas.	Sesgos: Los LLMs, entrenados con grandes cantidades de datos, pueden contener distorsiones o inclinaciones particulares que pueden ser reproducidos y potencialmente amplificados.
Gobernanza, ética y aspectos legales	Atención al cliente mejorada: Los LLMs impulsan chatbots sofisticados que ofrecen respuestas rápidas y sencillas a los consumidores y asumen funciones de atención al cliente.	Seguridad y Privacidad: La capacidad de la IA para recopilar grandes volúmenes de datos plantea preocupaciones de privacidad y su interconexión con otras aplicaciones o LLMs, temas de control de acceso y de integridad de los modelos.
Social y laboral	Mejora del talento: La IA puede ayudar a mejorar ampliamente los trabajos de millones de personas a nivel global, para lograr sus resultados en el menor tiempo.	Desplazamiento Laboral: La automatización puede causar que entre 400 y 800 millones de personas tengan que cambiar de trabajo para el año 2030. Esto requiere nuevas iniciativas para reeducar y volver a formar al personal (Reskilling y Upskilling)

Nota: Elaboración propia basada en Rouhiainen, 2018

Es por esto que esta edición de la revista *Sistemas*, de la Asociación Colombiana de Ingenieros de Sistemas −ACIS−, apunta a revisar, explorar y analizar los retos y oportunidades de la inteligencia artificial generativa, con el fin de traer al escenario actual diferentes posturas sobre el tema, como insumo para plantear alternativas y opciones en un entorno de disrupción tecnológica acelerada. Con ese propósito fueron convocados profesionales especialistas, quienes desde su experiencia proponen reflexiones para seguirle la pista al desarrollo de las tendencias y prácticas vigentes en esta temática, capitalizando lecciones aprendidas, repensando las dinámicas de los negocios y retos actuales, así mismo explorar el futuro que se avizora en el horizonte.

La columna de esta edición estuvo a cargo del señor Héctor Calderazzi, Vicepresidente del Capítulo de ISACA, Buenos Aires, Argentina, quien desarrolla una perspectiva analítica sobre la inteligencia artificial en la oscuridad (*Shadow AI*) de forma constructiva, argumentando que la respuesta no es prohibir, sino habilitar, para lo cual revisa lecciones históricas de riesgos semejantes en el pasado y así, justificar la necesidad de un enfoque proactivo con el fin de plantear interrogantes y reflexiones que motiven una vista enriquecida sobre las implementaciones y usos de proyectos con inteligencia artificial generativa.

La entrevista efectuada a la Dra. Mariana Sánchez Caparrós, académica de la Universidad de Buenos Aires, Argentina, detalla los retos alrededor de cómo va a cambiar la dinámica de las organizaciones en el mediano y largo plazo con la incorporación de la IA Gen, qué precauciones tomar y cómo sacarle todo su potencial y cómo comunicar su promesa de valor a una junta directiva.

La investigación realizada por los ingenieros Jeimy J. Cano M. y Andrés R. Almanza J., hace una síntesis del documento sobre la preparación de Colombia para incorporar la inteligencia artificial elaborada en conjunto con la UNESCO y el gobierno nacional, donde se detalla el diagnóstico general de la situación en Colombia y las recomendaciones concretas para avanzar en este proceso.

Se presenta un primer artículo realizado por los académicos Andrés Aguilera (Escuela de Negocios del Tecnológico de Monterrey, Campus Puebla, México) y Carolina Saldaña (Facultad de Administración, Universidad Externado de Colombia), quienes documentan un ejercicio de titulado "*Text to Anything*", llevado a cabo en un ambiente universitario (nivel pregrado y posgrado) en ciencias empresariales, donde los estudiantes de manera guiada experimentaron con diversas herramientas y casos de uso de la inteligencia artificial para la crea-

ción de diferentes tipos de contenidos (texto, código, audio, video) a partir de indicaciones en lenguaje natural, mejor conocidos como: *prompts*.

Un segundo artículo desarrollado por el ingeniero Jeimy J. Cano M., donde se hace un introducción al riesgo algorítmico, un riesgo emergente definido como la posibilidad de daño, pérdida financiera o afectación de la reputación empresarial que surge del uso, despliegue o explotación maliciosa de sistemas de IA, que hace evidente las limitaciones propias de los modelos tradicionales de seguridad y control.

Un tercer artículo desarrollado por el ingeniero Joshsua González, donde se analizan la relación entre las redes neuronales biológicas y las redes neuronales artificiales. Este artículo examina los principios fundamentales de cada enfoque y analiza sus puentes conceptuales, sin reducir uno en el otro, mostrando que la interacción entre neurociencia e inteligencia artificial no persigue imitación estructural, sino inspiración funcional y expansión mutua.

Un cuarto artículo desarrollado por el ingeniero Rafael González, profesor titular de la Pontificia Universidad Javeriana, que analiza la burbuja especulativa de la inteligencia artificial generativa como un fenómeno similar de la burbuja punto-com, pero que en la actualidad pueden ser más rápidas, simultáneas o extendidas. El reto es aprovechar el momento para tener una mira estratégica para concretar oportunidades en el ecosistema digital y la capacidad para asumir riesgos de forma equilibrada y así, innovar con éxito.

Un quinto artículo desarrollado por el ingeniero Julio López, consultor experto en proyectos y en inteligencia artificial, que presentan al lector no técnico, algunos de los mitos actuales alrededor de la Inteligencia Artificial (IA) incluyendo algunos conceptos para aclararlos, con el fin de entender las novedades tecnológicas y evaluar así su mejor utilización.

Un sexto artículo presentado por el ingeniero Juan Mario Posada, consultor senior de Accenture, desarrolla la dualidad del uso de la inteligencia artificial generativa para defensa y ataque bajo la óptica del NIST Cybersecurity Framework 2.0 con énfasis en la función de gobierno como base de la gobernanza algorítmica. Asimismo, examina el uso ético de esta tecnología en la academia y la empresa, junto con su impacto en la brecha digital global.

El foro de esta revista fue un espacio para compartir visiones desde diferentes ángulos sobre la inteligencia artificial generativa. Los ingenieros Germán Noguera (ACIEM – Asociación Colombiana de Ingenieros) y Fabio Rojas (KPMG), en conjunto con el contador público

Guillermo Zegarra (Pacífico Seguros, Perú), desarrollaron un diálogo abierto y nutrido para contrastar y complementar sus posturas alrededor de los retos que implica incorporar la IA Gen en las organizaciones. Ellos advierten sobre la necesidad promover una cultura de experimentación y pruebas permanente, para aprender rápido y asegurar que los resultados que se generan son los esperados. Igualmente advierten sobre las tensiones que se revelan con ocasión de eliminación de cargos y creación de nuevos por cuenta de la IA, sin dejar de lado los aspectos éticos claves para una implementación segura y responsable.

En resumen, se trata de un panorama renovado y provocador de nuevas transformaciones, retos y propuestas alrededor de la inteligencia artificial generativa, que tensionan las certezas de los saberes y prácticas existentes a la fecha en las realidades y perspectivas de las empresas actuales. Su contenido invita a todos los profesionales en las diferentes áreas del conocimiento a explorar las nuevas realidades de un mundo digital, tecnológicamente modificado y ahora habilitado por la inteligencia artificial, sin perjuicio de los nuevos desafíos políticos, económicos, sociales, tecnológicos, legales y ecológicos, donde la IA plantea, revela y reescribe nuevas incertidumbres y potencia el desarrollo de capacidades de negocio antes inexistentes, de cara a los riesgos que permanecen ocultos y latentes en los ahora ecosistemas digitales inteligentes de negocio.

Referencias

Alammar, J., & Grootendorst, M. (2024). *Hands-on large language models: Language understanding and generation*. O'Reilly Media.

McCowan, S., Krumbmüller, F. & Jaggi, G. (2025). *How can reimagining risk prepare you for an unpredictable world?* EY Insights. https://www.ey.com/en_us/insights/consulting/how-can-reimagining-risk-prepare-you-for-an-unpredictable-world

Rouhiainen, L. (2018). *Inteligencia artificial. 101 cosas que debes saber hoy sobre nuestro futuro*. Barcelona, España: Editorial Planeta.

***Jeimy J. Cano M., Ph.D, CFE, CICA**. Ingeniero y Magíster en Ingeniería de Sistemas y Computación por la Universidad de los Andes. Especialista en Derecho Disciplinario por la Universidad Externado de Colombia. Ph.D en Business Administration por Newport University, CA. USA. y Ph.D en Educación por la 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–.*

ACIS
ASOCIACIÓN COLOMBIANA DE INFORMÁTICA, SISTEMAS Y TECNOLOGÍAS AFINES

2026

Calendario de eventos

FEBRERO

CURSO VIRTUAL: DESARROLLO Y ARQUITECTURA DE SOFTWARE IA

13 FEB

ANIVERSARIO 50 AÑOS DE ACIS.

17 MAR

ASAMBLEA ACIS

4 - 8 MAR

III PROGRAMADORES DE AMÉRICA 2026 (SANTIAGO DE CHILE - CHILE)

ABRIL

JORNADO DE GESTION DE PRODUCTOS Y PROYECTOS TI

MAYO

CURSO VIRTUAL: CALIDAD DE SOFTWARE CON IA

10, 11 Y 12 JUN

ENCUENTRO REDIS 10, 11 Y 12 CHINÁCOTA NORTE DE SANTANDER

JUNIO

SIMPOSIO GESTIÓN DE DATOS

28, 29 Y 30 JUL

JORNADA INTERNACIONAL DE SEGURIDAD INFORMÁTICA.

AGOSTO

ACISTIC 2026

SEPTIEMBRE

JORANDA GEODATOS

13 - 17 OCT

ENCUENTRO REDUC@TE 13 AL 17

OCTUBRE

MARATÓN NACIONAL DE PROGRAMACIÓN 2026

NOVIEMBRE

MARATÓN REGIONAL LATINOAMERICANA DE PROGRAMACIÓN 2026

DICIEMBRE

CURSO VIRTUAL: INTERPRETACIÓN DE LAS IMAGENES SATELITALES

MÁS INFORMACIÓN

301 5530540

WWW.ACIS.ORG.CO

Shadow AI: Riesgo emergente y habilitación

DOI: 10.29236/sistemas.n177a2

Resumen

La Inteligencia Artificial Generativa ha emergido como un catalizador de cambio, entregando agilidad al negocio, pero planteando un desafío fundamental: gobernarla con propósito y confianza. Esta democratización impulsa a los usuarios a convertirse en desarrolladores, lo que genera la *Shadow AI* (TI en la Sombra), donde las soluciones se implementan fuera del marco de gobernanza por la necesidad de velocidad. El artículo examina la *Shadow AI* de forma constructiva, argumentando que la respuesta no es prohibir, sino habilitar. Se revisan lecciones históricas de riesgos (ej. hojas de cálculo) para justificar la necesidad de un enfoque proactivo de la función de Riesgos, que debe actuar como facilitador seguro, según el Dominio 2 (ISACA CRISC, 2024), superando actitudes conservadoras que imponen controles desalineados con el costo/beneficio. Se destaca la importancia de integrar a los desarrolladores en el proceso de seguridad, mientras sus roles se transforman de la codificación rutinaria hacia la gestión de infraestructura de IA y la arquitectura de *prompts*, requiriendo recapacitación. Se concluye que la IA debe ser un facilitador del negocio, requiriendo un paraguas de gobierno centralizado para canalizar la innovación de la sombra hacia una adopción estratégica y segura. Herramientas como el CASB (*Cloud Access Security Broker*) son esenciales para lograr la visibilidad y el control efectivo en entornos *cloud*.

Palabras Claves

IA Generativa, Shadow AI, Gobierno de la IA, Riesgo Tecnológico, Habilitación de Negocio, CASB

Héctor Calderazzi

Introducción

La Inteligencia Artificial Generativa no es una simple herramienta de productividad; es un catalizador de cambio que está entregando un poder significativo al usuario de negocio, prometiendo una agilidad sin precedentes. Esta democratización, sin una adecuada canalización, genera una fricción natural con los principios de gobernanza, seguridad y ética, impactando la Confianza Digital de la organización (Moyle, 2023).

Esto me alertó sobre el escenario de riesgo creciente: la Shadow AI, que es la manifestación más reciente del concepto amplio de TI en la Sombra (Shadow TI), definido como el uso de sistemas, *software* y servicios por fuera del circuito establecido para gobernar adecuadamente el ciclo de vida de los sistemas de información y sin considerar las áreas de TI y Seguridad de la Información entre otras.

Exploraremos las motivaciones positivas detrás de la Shadow AI, analizaremos cómo las lecciones del pasado y el rol proactivo del profesional de riesgos (CRISC – *Certified in Risk and Information Systems Control*) nos obligan a buscar soluciones de habilitación en lugar de prohibición. El objetivo es claro: demostrar cómo la IA debe ser un facilitador del negocio, guiando su adopción con propósito y seguridad.

La búsqueda de agilidad y el origen de la sombra

Se observa una tendencia clara: los usuarios quieren respuestas, no datos.

Además, las propias empresas de *software*, con una estrategia pujante, vienen en forma entusiasta por más y empujan a que cada usuario sea un desarrollador. Esta combinación de demanda de agilidad y la facilidad para acceder a *software* con capacidades de IA entusiasma a los usuarios y da origen a la Shadow AI, definida por ISACA como la “Nueva Frontera

del Riesgo Empresarial" (Rajasekharan, 2025), donde la proliferación de servicios de IA en aplicaciones comerciales está agravando el uso en la sombra (Moyle, 2023). El principal riesgo es que estas soluciones operan fuera del marco de gobernanza, arriesgando la Fuga de Propiedad Intelectual, el Sesgo y el Incumplimiento Normativo.

La Shadow TI (y su evolución, la Shadow AI) no surge por intención maliciosa, sino de la necesidad urgente de velocidad del negocio.

Los empleados y gerentes implementan estas soluciones *ad hoc* porque:

- Velocidad vs. Proceso: El proceso formal de TI para aprobar herramientas es percibido como lento y burocrático.

- Funcionalidad: Las herramientas sancionadas por TI no satisfacen las necesidades específicas.

- Descentralización: La facilidad para acceder a servicios *cloud* permite a las unidades de negocio resolver sus problemas con presupuestos departamentales.

Lecciones aprendidas: análisis de patrones de riesgo histórico

Para gestionar la Shadow AI de forma constructiva, es esencial analizar el patrón de comportamiento histórico de riesgo y aprender del patrón de riesgo tecnológico pasado. La historia de TI nos enseña que la Shadow TI aparece cuando la agilidad del negocio supera la capacidad de respuesta de la gobernanza, independientemente de la tecnología en uso.

- El riesgo de las hojas de cálculo masivas:

- Tecnología: Masificación de herramientas como Excel, utilizadas como bases de datos y motores de cálculo críticos (Aplicaciones de Usuario Final - EUC). Este riesgo sigue vigente hoy.

- Riesgo: Las planillas se democratizaron sin controles de identificación, trazabilidad o integridad (Tsang et al., 2022; Rodríguez, 2019). Esto genera errores de formulación que impactan decisiones financieras y comerciales (EUSPRIG, s.f.). El riesgo aquí no era el *software* en sí, sino la ausencia de un marco de gobernanza sobre información crítica fuera de los sistemas corporativos.

- Descentralización de datos y extracción (ETL):

- Tecnología: Uso creciente de herramientas de extracción de datos y transformación de datos (similares a ETL - Extracción, Transformación y Carga o *reporting tools*). El objetivo primario fue que cada área generase sus propios reportes.

- Riesgo: La dificultad de mantener y evolucionar los procesos ETL (que requieren cambios constantes de código debido a variaciones en los sistemas de origen) genera un alto riesgo de inconsistencia de los datos (Reed et al., 2010). Por ejemplo, errores en la tabla de tasas de comisiones de ventas pueden resultar en un cálculo erróneo del importe de la comisión.

  Al descentralizar este acceso sin el *expertise* de TI, la empresa sufre por la falta de integridad y la aparición de múltiples versiones de la verdad en la reportería.

• Interacción directa del usuario con el proveedor externo:

- Tecnología: La interacción de las unidades de negocio con el proveedor externo para la adquisición directa de *software*, servicios o la contratación de desarrolladores, evitando los canales formalmente establecidos por TI.

- Riesgo: Evitar el CVDS (Ciclo de Vida de Desarrollo de Sistemas) genera debilidad en el control de cambios y pérdida de documentación (IBM, s.f.), comprometiendo la seguridad, la continuidad operativa y la sostenibilidad del sistema.

La Shadow AI amplifica estos desafíos. El riesgo ahora no es una fórmula incorrecta, sino una decisión algorítmica no validada. La lección es clara: la solución nunca ha sido la prohibición, sino la habilitación controlada. El profesional de riesgos debe ser diligente en la identificación de estas soluciones en la sombra, estudiando los flujos de trabajo reales y revisando *logs* para entender las conexiones no aprobadas, aceptando que la Shadow TI existe, según el Dominio 2 (ISACA CRISC, 2024).

La postura conservadora vs. la habilitación proactiva

La forma en que se enfrenta la Shadow AI distingue a las organizaciones. Por un lado, tenemos la actitud basada en la seguridad pura, y por el otro, la que busca la habilitación:

• Caso de la actitud conservadora (Las 20 Observaciones):
  Recientemente, un colega me compartió un caso muy ilustrativo: el equipo de TI había identificado la solución ágil a una necesidad operativa de negocio mediante la implementación de un agente de IA. Sin embargo, la respuesta del equipo de control fue una lista de 20 observaciones que, si bien estaban técnicamente justificadas, no habían sido calibradas con una evaluación del impacto de los riesgos al negocio. Este enfoque, centrado solo en el riesgo inherente, implicó que la lista de 20 controles resultara más cara de implementar que el beneficio que la nueva aplicación aportaría al negocio.

Esta desalineación entre el control y la oportunidad frenó la innovación y refuerza la percepción del área de control como un obstáculo burocrático, sin que la intención fuera detener el negocio.

- El CISO como habilitador:
  Este caso ilustra por qué el Dominio 4 (ISACA CRISC, 2024) es vital: el profesional de riesgos debe ser un facilitador seguro que equilibre los controles con el valor. Si TI es reticente, el negocio encontrará su propia solución. El rol es liderar el esfuerzo para demostrar cómo se pueden incorporar nuevas tecnologías de forma segura en lugar de rechazar su adopción.

- **Habilitación controlada y el *sandbox*:** La solución es proporcionar entornos *sandbox* (cajas de arena) para la **experimentación segura** (Ramachandran, 2025). Esto permite a los usuarios innovar y luego, si el piloto demuestra valor, se inicia un proceso de cambio controlado donde TI profesionaliza la solución.

- **Riesgo Cognitivo:** La habilitación incluye abordar el **riesgo cognitivo** (Kos'myna, 2025), fortaleciendo el **criterio humano** y el **pensamiento crítico** para evitar la **aceptación pasiva** de las respuestas de la IA.

Gobierno de la IA: hacia un modelo centralizado de facilitación

La gestión de la **Shadow AI** es una elección cultural que define la **postura de riesgo** de la organización, como se observa en la tabla No.1.

El desafío es encontrar el nivel de gris que equilibre la necesidad de competir con la responsabilidad correspondiente.

La Shadow TI es catalogada como una fuente clave de Riesgo Emergente, según el Dominio 3 (ISACA CRISC, 2024) debido al Riesgo Fuera de Apertura y el Impacto Normativo. La mitigación se basa en la

Tabla 1

Postura de riesgo de la organización frente a la Shadow AI

Actitud	Riesgo asumido	Innovación	Resultado
Conservadora	Bajo	Bajo	Seguridad y control, pero rezago competitivo.
Arriesgada	Alto	Alto	Innovación y liderazgo, pero alta exposición a riesgos (regulatorios, éticos, operacionales).

Nota: Elaboración propia.

visibilidad y la aplicación de políticas, lo que requiere el uso de herramientas de control adecuadas, según el Dominio 4 (ISACA CRISC, 2024). Entre estas herramientas para gestionar el riesgo de la sombra se encuentran las soluciones de Prevención de Pérdida de Datos (DLP), los sistemas de Monitoreo de Eventos e Información de Seguridad (SIEM) y, fundamentalmente, las herramientas especializadas en entornos *cloud*.

- El CASB como Puerta de Enlace de Seguridad:

  Para lograr la visibilidad y el control necesarios sobre las aplicaciones en la sombra, las herramientas CASB (*Cloud Access Security Broker* o Agente de Seguridad de Acceso a la Nube) son esenciales (Gartner, 2025). Un CASB actúa como un punto de aplicación de políticas de seguridad situado entre los usuarios y los proveedores de servicios cloud. Su función principal es descubrir las aplicaciones no sancionadas (Shadow TI), monitorear el comportamiento de los usuarios, aplicar políticas de DLP y garantizar el cumplimiento normativo en tiempo real.

- El Riesgo de Desplazamiento y la Integración del Desarrollador como Facilitador: Este cambio cultural implica la redefinición de roles en TI. El desafío clave es sumar a los desarrolladores como facilitadores de soluciones seguras para que acompañen en la concienciación de los usuarios e incluso brinden su apoyo para que los usuarios logren mejores soluciones, justo cuando enfrentan la incertidumbre de cómo evolucionarán sus funciones en el contexto de la IA.

- Pruebas Rigurosas: La guía de ISACA Emerging Technology (2024), referente a la Ley de IA de la UE, enfatiza la necesidad de *testing*, documentación técnica y la creación de *sandboxes* regulatorios. Ahora bien, es el ISACA CRISC (2024), a través de su *Review Manual*, el que profundiza en los requisitos de control y aseguramiento que debe aplicar el profesional de riesgos. Dicho manual detalla la metodología para la validación de modelos críticos (incluyendo lotes de prueba y análisis de resultados) y se alinea con las directrices regulatorias aplicables a modelos de riesgo financiero.

- La Calidad Continua (Desarrollo Futuro): Para garantizar que la solución se mantenga confiable, segura y sin sesgos en el tiempo, es crucial considerar disciplinas de calidad continua. Este tema es amplio y estratégico, y podría ser objeto de un desarrollo en detalle en otra columna en el futuro.

La respuesta a la Shadow AI impulsa una transformación radical en el ecosistema de roles de TI. La

codificación rutinaria migra hacia la gestión de infraestructura de IA y la función crítica de arquitecto de *prompts*, exigiendo re-capacitación general para profesionalizar soluciones y dar soporte a los usuarios que explotarán las funciones y beneficios de AI. El factor humano se confirma como insustituible: el análisis de exposición laboral (el 'iceberg' de Chopra et al., 2025) muestra que el juicio experto, la ética y la supervisión crítica no son automatizables. Precisamente, el desafío fundamental del profesional de riesgos es actuar como facilitador seguro que equilibre los controles con el valor, según el Dominio 4 (ISACA CRISC, 2024). La IA no se protege a sí misma; es el talento humano quien, con criterio y visión, debe liderar su desarrollo seguro, calidad y cumplimiento continuo (Rajasekharan, 2025).

Referencias

Chopra, A., Bhattacharya, S., Salvador, D., Paul, A., Wright, T., Garg, A., Ahmad, F., Schwarze, A. C., Raskar, R., & Balaprakash, P. (2025). *The Iceberg Index: Measuring Workforce Exposure Across the AI Economy*. arXiv. https://arxiv.org/abs/2510.25137

European Spreadsheet Risks Interest Group (EUSPRIG). (s.f.). *Investigación sobre errores en hojas de cálculo y sus consecuencias.* https://eusprig.org/research-info/horror-stories/

Gartner. (2025). *Definition: Cloud Access Security Broker (CASB)*. https://www.gartner.com/en/information-technology/glossary/cloud-access-security-brokers-casbs

IBM. (s.f.). *What Is Shadow IT?* https://www.ibm.com/think/topics/shadow-it#:~:text=Shadow%20IT%20is%20any%20software,department's%20approval%2C%20knowledge%20or%20oversight

ISACA CRISC. (2024). *CRISC Review Manual* (7th ed.).

ISACA Emerging Technology. (2024). *Understanding the EU AI Act: Requirements and Next Steps*. https://www.compliancehub.wiki/content/files/2024/10/ISACA_Understanding_EU-AI-Act.pdf

Kos'myna, N. (2025). *Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task*. MIT Media Lab. https://www.media.mit.edu/publications/your-brain-on-chatgpt/

Moyle, E. (2023). *Digital Trust and Adopting Generative AI*. ISACA Journal, 5. https://www.isaca.org/resources/isaca-journal/issues/2023/volume-5/digital-trust-and-adopting-generative-ai

Rajasekharan KR, CISM, CDPSE, PMP. (2025). *From Shadow IT to Shadow AI: Navigating the New Frontier of Enterprise Risk*. ISACA Newsletters, 19. ISACA. https://www.isaca.org/resources/news-and-trends/newsletters/atisaca/2025/volume-19/from-shadow-it-to-shadow-ai-navigating-the-new-frontier-of-enterprise-risk

Ramachandran, R. (2025). *Safeguarding the Future: Strategies for Protecting Generative AI, LLMs, and Agentic AI*. ISACA.

https://www.isaca.org/resources/news-and-trends/isaca-now-blog/2025/safeguarding-the-future-strategies-for-protecting-generative-ai-llms-and-agentic-ai

Reed, C., Wang, Y., & Dutta, A. (2010). *Achieving Data Warehouse Nirvana*. ISACA Journal, 4. https://www.isaca.org/-/media/files/isacadp/project/isaca/articles/journal/archives/journal-volume-4-2010.pdf

Rodríguez, I., C.P. (Auditor y Consultor, Diplomado en Alta Gerencia de Seguros, Especialista en Dirección Financiera y Desarrollo Organizacional). (2019). *Las hojas de cálculo y los riesgos para el Auditor*. Auditool. https://www.auditool.org/blog/auditoria-externa/las-hojas-de-calculo-y-los-riesgos-para-el-auditor

Tsang, B., Ward, S., Zhang, L., & Storey, M. (2022). *Our Approach Managing Risk of End User Computing (EUC)*. KPMG International. https://assets.kpmg.com/content/dam/kpmgsites/uk/pdf/2022/10/kpmg-euc-proposition-sep-2022.pdf

Héctor Calderazzi, CISA, CRISC, CISM
Profesional de Tecnología de la Información (TI) con más de 45 años de trayectoria, cuyo expertise se ha desarrollado principalmente en entidades financieras, especializándose en Seguridad de la Información, Auditoría de Sistemas y Gestión de Riesgos (GRC). Posee un Postgrado en Innovación Empresarial (UCEMA) y una Diplomatura en Gobernanza de Datos. Es un profesional certificado por ISACA con las credenciales CISA (Certified Information Systems Auditor), CRISC (Certified in Risk and Information Systems Control) y CISM (Certified Information Security Manager). En su experiencia, lideró la seguridad y la gestión de riesgos en el sector financiero, destacando su participación en fusiones institucionales y proyectos estratégicos de alta complejidad. Actualmente, combina la consultoría senior en Gobierno de TI, Gestión de Riesgos y Seguridad de la Información con su rol de Mentor ISACA para capítulos a nivel global y Docente en diversas entidades académicas. Es vicepresidente de ISACA Buenos Aires Chapter.

Pormenores de la Inteligencia Artificial

Desde la óptica de Mariana Sánchez Caparrós.

DOI: 10.29236/sistemas.n177a3

"Mi trabajo busca impulsar la transformación digital responsable en el sector público. Dirijo equipos interdisciplinarios para diseñar políticas de gobernanza tecnológica y desarrollar soluciones éticas de IA, especialmente en los ámbitos judicial y administrativo", indica la entrevistada en este número de la revista Sistemas.

Mariana Sánchez Caparrós es doctora en Ciencias Jurídicas, magíster en Derecho Administrativo y abogada. Como subdirectora de IALAB, lidera investigaciones y proyectos en todos los asuntos relacionados con derecho, ética e inteligencia artificial, enfocada en la gobernanza de tecnologías emergentes y la protección de derechos fundamentales.

Tales actividades profesionales las combina a la perfección con el deporte, la lectura, la docencia, las caminatas y el trabajo creativo con equipos interdisciplinarios. "Me interesan especialmente las películas y libros que plantean dilemas humanos frente a la tecnología como Ready Player One", señala.

Revista Sistemas
¿Cómo va a cambiar la IA generativa la dinámica de las organizaciones en el mediano y largo plazo?

Mariana Sánchez Caparrós
La IA generativa está obligando a repensar la arquitectura completa del trabajo; no solo automatiza tareas, sino que redefine procesos, roles y formas de crear valor.

En el mediano plazo veremos organizaciones híbridas, donde el flujo de trabajo combina agentes inteligentes, automatizaciones avanzadas y supervisión humana significativa.

En el largo plazo, la ventaja competitiva no estará en “usar IA”, sino en gobernarla bien; en otras palabras, tener claridad estratégica, procesos auditables, criterios éticos explícitos y capacidades internas para diseñar, validar y monitorear sistemas complejos. Así mismo, en elegir cuándo, cómo y qué

IA utilizar, en diseñar modelos propios o integrar modelos desarrollados por terceros, según el caso de uso.

Las organizaciones que no logren integrar esta mirada sociotécnica quedarán fuera de los estándares emergentes globales.

RS: *¿Qué hace tan seductor incorporar IA Gen hoy? ¿Qué precauciones tomar y cómo aprovechar su potencial?*

MSC: La seducción viene de su inmediatez; permite acelerar decisiones, producir contenido, analizar grandes volúmenes de información y agilizar procesos que antes demandaban meses. Sin embargo, el entusiasmo no puede alejarse de la responsabilidad.

La IA generativa requiere marcos de gobernanza claros, segmentación de riesgos, evaluaciones de impacto, validación humana y criterios sólidos de privacidad. Su verdadero potencial aparece cuando se la integra a procesos bien diseñados, evitando la "IA-por-IA" y priorizando casos de uso con impacto, trazabilidad y valor público o corporativo verificable. El punto es experimentar para construir capacidades sostenibles.

RS: *¿Cómo llevar el mensaje a una junta directiva? ¿Cómo comunicar la nueva promesa de valor?*

MSC: Los directorios necesitan entender que la IA generativa no es un gasto tecnológico, sino un activo estratégico. Se comunica mejor cuando se articula en tres planos:

1. Valor: reducción de tiempos, eficiencia operativa, nuevas capacidades analíticas, mejoras en la experiencia del cliente o del ciudadano.

2. Riesgo: qué controles, auditorías, gobernanza y resguardos legales se prevén. La confianza es un habilitador del negocio y nos permite cuidar la marca.

3. Hoja de ruta: pasos concretos, responsables, métricas de éxito y un modelo de adopción progresiva.

El lenguaje debe ser claro, sin tecnicismos innecesarios, y respaldado por pilotos que demuestren impacto real.

RS: *¿Es necesario auditar modelos antes de desplegarlos? ¿Auditar código para asegurar desempeño y resultados?*

MSC: Depende del tipo de solución de IA. En soluciones que integran modelos generativos de terceros o se apoyan en plataformas de terceros, las organizaciones no pueden auditar el código, pero sí deben auditar sus usos: prompts, flujos internos, riesgos, datos involucrados y validaciones humanas.

Cuando se trata de modelos propios, de gobierno o empresariales, sí corresponde revisar componentes críticos del pipeline: datos de entrenamiento, sesgos, métricas de desempeño, condiciones de despliegue y mecanismos de rendición de cuentas.

En todos los casos, más que una auditoría puntual, se necesita un ciclo de gobernanza continua, con monitoreo, logging, explicabilidad funcional o contextual, transparencia proporcional al caso de uso e impacto en derechos, y actualizaciones controladas.

RS: *¿En la era de la IA, ¿la confianza digital es el valor más importante que una empresa puede ofrecer?*

MSC: Sin duda. La confianza digital se transformó en un activo económico y reputacional. Las personas necesitan saber que sus datos están protegidos, que los sistemas son auditables y que la tecnología no afecta derechos.

La confianza se construye combinando infraestructura segura, buenas prácticas de privacidad, transparencia, uso proporcional, explicabilidad y una cultura organizacional que priorice la ética y la interdisciplina como parte del diseño.

En un entorno de creciente complejidad tecnológica, la confianza es un atributo valorado por clientes y socios comerciales.

Diagnóstico de la preparación de Colombia para la implementación de la recomendación de la UNESCO sobre la Ética de la Inteligencia Artificial (IA)

DOI: 10.29236/sistemas.n177a4

Jeimy J. Cano M.

Andrés R. Almanza J.

Resumen

Este artículo presenta un resumen de los resultados de la aplicación de la Metodología de Evaluación del Estadio de Preparación (RAM) de la UNESCO para evaluar la capacidad de Colombia para implementar la Recomendación sobre la Ética de la Inteligencia Artificial (IA). El análisis por dimensiones revela fortalezas importantes en el marco jurídico, y notables desarrollos en los sistemas estadísticos a cargo del DANE (Departamento Administrativo Nacional de Estadística). Sin embargo, persisten desafíos estructurales significativos: el gasto en Investigación y Desarrollo (I+D+i) es bajo, limitando la generación de soluciones propias. También existen importantes brechas digitales a nivel territorial (urbano-rural) y de género en la formación CTIM (Ciencia, Tecnología, Ingeniería y Matemáticas). Las principales recomendaciones de política se centran en el aumento de la inversión en I+D+i y el fortalecimiento de la gobernanza interinstitucional y la supervisión de riesgos en tecnologías de IA a nivel sectorial.

Palabras clave

Ética, I+D+i, Gobernanza, Brechas digitales, RAM UNESCO

Introducción

En noviembre de 2021, los 193 Estados miembros de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) adoptaron la *Recomendación sobre la Ética de la Inteligencia Artificial*, el primer instrumento normativo global sobre este tema. Este marco busca asegurar el desarrollo y uso ético de la IA (Inteligencia Artificial), abarcando derechos humanos, dignidad humana, y sostenibilidad ambiental, traducidos en principios como la rendición de cuentas, la transparencia y la privacidad.

La evaluación realizada aplica la Metodología de Evaluación del Estadio de Preparación (RAM) de la UNESCO (2025), una herramienta diagnóstica diseñada para ayudar a los Estados miembros, como Colombia, a entender su nivel de preparación institucional, regulatoria, y de datos para implementar la recomendación de manera ética y responsable, que se complementa con diagnósticos recientes, como la Evaluación del Panorama de la Inteligencia Artificial (AILA) (PNUD, 2024). La RAM analiza cinco dimensiones clave: jurídica, social y cultural, científica y educativa, económica, y técnica y de infraestructura.

Colombia ha demostrado un compromiso progresivo y una visión fundamentalmente optimista respecto a la gobernanza ética de la IA, considerándola compatible y esencial para la innovación y el crecimiento económico. Según el Índice Latinoamericano de Inteligencia Artificial (ILIA) 2025. Colombia ha sido clasificado como uno de los países con adopción moderada de IA (66.06 puntos en el subdimensión de Adopción), y se encuentra en el cuarto lugar en la dimensión de Investigación, Desarrollo y Adopción (I+D+A) con 48.6 puntos, solo detrás de Chile, Brasil y México (CEPAL, 2025).

Este ejercicio de diagnóstico se materializó con la colaboración activa de la Presidencia de la República y entidades clave como el Ministerio de Ciencia, Tecnología e Innovación (MinCiencias), el Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC), y el Departamento Nacional de Planeación (DNP)

Por tanto, este documento hace un resumen de este reporte donde se indican las fortalezas normativas del país, así como los desafíos persistentes en inversión en Investigación y Desarrollo (I+D), la brecha digital urbano-rural, y la fragmentación en la gobernanza de la IA.

Finalmente, se condensan las recomendaciones políticas específicas para el gobierno colombiano con el fin de fomentar un ecosistema de IA ético, responsable e inclusivo, alineado con el marco global de la UNESCO (2025).

Contexto de la política pública nacional

Colombia ha avanzado en el desarrollo de políticas públicas digitales y de IA desde hace más de siete años. Este esfuerzo se ha consolidado a través de documentos programáticos clave que delinean la hoja de ruta del país:

- CONPES 3920 de 2018: "Política Nacional de Explotación de Datos (Big Data)".

- CONPES 3975 de 2019: "Política Nacional para la Transformación Digital y la IA", considerada la primera política pública de IA, buscando condiciones habilitantes y adoptando el Marco Ético para la IA en Colombia (2021).

- Hoja de Ruta para el Desarrollo y Adopción de Inteligencia Artificial en Colombia (2024), liderada por MinCiencias, que promueve la adopción ética y sostenible a través de ejes como Ética, Adopción y Sostenibilidad.

- Estrategia "Colombia PotencIA Digital" (MinTIC): Lanzada en 2023, busca articular actores del ecosistema digital en torno a Conectividad, Educación Digital y Ecosistemas de Innovación.

- CONPES 4144 de 2025: "Política Nacional de Inteligencia Artificial", la más reciente y comprehensiva, establece una política pública en torno a seis ejes fundamentales: 1) Ética y gobernanza; 2) Datos e infraestructura; 3) Investigación, Desarrollo e Innovación (I+D+i); 4) Desarrollo de capacidades y talento digital; 5) Mitigación de riesgos; y 6) Uso y adopción de la IA. Este documento 2025 proyecta 106 acciones hasta 2030, con una inversión total aproximada de 479.273 millones de pesos colombianos (aproximadamente 117 millones de USD), lideradas por el DAPRE (Departamento Administrativo de Presidencia de la República), MinTIC, MinCiencias y DNP.

Diagnóstico de preparación nacional: fortalezas y desafíos clave

El diagnóstico de preparación nacional (RAM) identificó las fortalezas existentes en Colombia, como su normatividad en privacidad y acceso a la información pública, y el desarrollo de sistemas estadísticos. Sin embargo, también reveló brechas críticas que deben abordarse para asegurar la implementación ética de la Recomendación.

A. Dimensión Jurídica

Fortalezas:

- Marco Normativo de Datos Personales: Colombia cuenta con un Régimen de Protección de Datos Personales robusto (Leyes 1581 de 2012 y 1266 de 2008) que garantiza el derecho al hábeas data, alineado con prin-

cipios de legalidad, finalidad, libertad, transparencia y seguridad. La Superintendencia de Industria y Comercio (SIC) es la autoridad de control y ha desarrollado iniciativas exploratorias como el "Sandbox regulatorio sobre privacidad desde el diseño y por defecto en proyectos de IA".

- Acceso a la Información: La Ley 1712 de 2014 garantiza el derecho de acceso a la información pública bajo el principio de máxima publicidad.

- Desarrollo de Ciberseguridad: Colombia se encuentra en el nivel 3 de compromiso de ciberseguridad (T3: Establecimiento) según el Índice Global de Ciberseguridad de la UIT, destacando en habilidades técnicas. Además, se expidió recientemente la Estrategia Nacional de Seguridad Digital (2025-2027) con un Plan de Acción de 29 puntos.

- A pesar de la fragmentación general en la gobernanza de la IA, Colombia destaca en la adopción de IA en el gobierno, logrando el puntaje máximo (100 puntos) en el subindicador de Uso de IA en participación ciudadana del ILIA 2025 (CEPAL, 2025).

Desafíos:

- Necesidad de actualización normativa: Se identifica la pertinencia de actualizar instrumentos normativos existentes (como la Ley 1581 de 2012) para responder a las realidades específicas de los sistemas de IA. Hay iniciativas en curso, como el Proyecto de Ley 274 de 2025 Cámara.

- Fragmentación y articulación: Históricamente, ha habido desafíos en la articulación de actores públicos y privados, lo que ha llevado a una fragmentación en la construcción de políticas y regulación de IA. Se destaca positivamente el trabajo reciente de la Mesa Interinstitucional de IA para superar estas dificultades.

- Adquisición pública: Colombia carece de normatividad específica para la adquisición de sistemas de IA, aunque existe un marco general para contratación estatal. El CONPES 4144 busca mitigar esta falencia mediante lineamientos jurídicos para la incorporación de IA en compras públicas.

- A pesar de la existencia de un Marco Normativo de Datos Personales robusto (Leyes 1581 de 2012 y 1266 de 2008) y un Marco Ético de IA en el sector público, la implementación de la ética es desigual. Las subdimensiones de Seguridad y Transparencia en el uso ético de la IA se encuentran en un nivel 'Sistemático'. Específicamente, Colombia carece de herramientas

prácticas para verificar la transparencia en el uso de IA y existe un déficit en la información pública sobre los sistemas automatizados utilizados en el sector público. Este riesgo se manifiesta en un puntaje bajo de 24.22 de 100 en el aspecto de Responsabilidad según el Global Index on Responsible AI (GIRAI), y solo 4.95 de 100 en evaluación de impacto (PNUD, 2024).

B. Dimensión Social y Cultural

Fortalezas:

- Cierre de brecha digital de género: Hay avances en el cierre de la brecha de género en el uso de Internet, pasando de 0.918 a 1 entre 2015 y 2025. Además, se han implementado iniciativas específicas, como la Ley 2337 de 2023 y la convocatoria "Orquídeas: Mujeres en Inteligencia Artificial".

- Alta participación electrónica: Colombia tiene altos puntajes en el Índice de Desarrollo Gubernamental Electrónico (EGDI, puesto 68/193) y el Índice de Participación Electrónica (EPI, puesto 46/193), superando los promedios mundial, regional y subregional en 2024.

- Consideraciones ambientales: La hoja de ruta de MinCiencias y el CONPES 4144 de 2025 abordan la sostenibilidad ambiental, destinando recursos y proponiendo el uso de IA para enfrentar riesgos como el cambio climático y mejorar la eficiencia energética.

Desafíos:

- Brechas persistentes: Persisten importantes brechas de género en áreas CTIM (Ciencia, Tecnología, Ingeniería y Matemáticas) (solo el 13.76% de los grados en CTIM son de mujeres, frente al 35.12% de hombres). La brecha urbano-rural en el uso de Internet sigue siendo significativa (82.6% urbano vs. 59.6% rural).

- Confianza en IA: Aunque el 78% de los colombianos confía en el sector tecnológico, solo el 46% confía en el sector de la IA (una diferencia de -32 puntos).

- Ausencia de marcos específicos culturales: Colombia no cuenta con políticas específicas relacionadas con el uso de IA para la preservación del patrimonio cultural y de las lenguas indígenas, aunque existen planes generales de apoyo a las TIC y a las lenguas nativas.

C. Dimensión Científica y Educativa

Fortalezas:

- Posicionamiento en I+D regional: Colombia supera el promedio regional en el Índice Latinoamericano de Inteligencia Artificial (ILIA) en el área de Investigación, Desarrollo y Adopción,

con un alto número de publicaciones en IA (52,34 puntos, 20 puntos por encima de la media).

- Oferta académica creciente: Existe un creciente interés y oferta de programas de posgrado (maestrías y especializaciones) en IA en diversas universidades. Se destaca la reciente creación de la primera Facultad de Inteligencia Artificial en el país (Universidad de Caldas).

- Iniciativas de capacitación pública: Programas como AvanzaTEC (MinTIC y Code.org) y SENATIC (SENA, MinTIC y OIT) ofrecen cursos gratuitos en IA y habilidades digitales, buscando la apropiación social y la inclusión de ética en la formación.

Desafíos:

- Bajo Gasto en I+D: El gasto en Investigación y Desarrollo (GERD) en Colombia es críticamente bajo, situándose en solo el 0.29% del PIB (2023), muy por debajo del promedio de la OCDE. El gobierno se ha fijado la meta de aumentarlo al 0.5% para 2026.

- Baja Innovación y Patentes: Entre 2019 y 2023, Colombia solo presentó una solicitud de patente en IA, contrastando fuertemente con otros países de la región (Brasil con 198, México con 106). A pesar de esta limitación en el gasto, la capacidad académica ha progresado: en la última versión del ILIA 2025 (CEPAL, 2025), Colombia fue uno de los cuatro países latinoamericanos que se incorporaron recientemente a la lista de naciones que ofrecen programas de doctorado en IA, duplicando la oferta regional. Además, el país es destacado por su desempeño en capacidad de Unidades de Procesamiento Gráfico (GPU) per cápita, junto a Uruguay y Costa Rica, lo que indica potencial para el cómputo avanzado en IA.

- Brecha de habilidades y currículos: Existe escasez de capital humano para el diseño, desarrollo y despliegue de sistemas de IA. Aunque hay una estrategia general de inclusión de tecnologías digitales (CONPES 3988 de 2020), es necesario detallar líneas de acción específicas para la IA. Además, el país tiene un nivel de habilidades en ciencia de datos significativamente por debajo del promedio global (24/100 en el Global Skills Report de Coursera).

- La subdimensión de habilidades para la gestión y uso de soluciones de IA en el sector público se encuentra en un nivel 'Sistemático' (2.8/5). Esta debilidad es percibida por los propios funcionarios, ya que casi el 100% de los participantes en la encuesta AILA reconocen que no existen suficientes capacidades técnicas avanzadas para construir

herramientas de IA para el sector público. Esta brecha de habilidades avanzadas es una oportunidad de mejora prioritaria (PNUD, 2024).

D. Dimensión Económica

Fortalezas:

- Crecimiento en empleos TIC: El número de personas ocupadas en ocupaciones especializadas en Tecnologías de la Información y las Comunicaciones (TIC) aumentó en cerca de 230,000 entre 2015 y 2023, alcanzando 448.026 personas en 2023.

- Enfoque en exportación de servicios: Hay un enfoque hacia la exportación de servicios de contenido tecnológico y TIC.

Desafíos:

- Escasez de talento humano: El CONPES 4144 de 2025 señala la escasez de capital humano para el diseño, desarrollo, despliegue, uso, monitoreo y evaluación de sistemas de IA. Se requiere fortalecer las habilidades TIC del capital humano nacional.

- Baja inversión: El gasto bruto en I+D sigue siendo bajo (0.2% del PIB).

E. Dimensión Técnica y de Infraestructura

Fortalezas:

- Desempeño estadístico sólido: El Departamento Administrativo Nacional de Estadística (DANE) lidera un sistema estadístico robusto, con un alto puntaje en los Indicadores de Desempeño Estadístico (SPI) del Banco Mundial (87.3 sobre 100). El sistema incluye un Sistema de Ética Estadística (SETE) y un Instrumento de Evaluación de la Calidad Estadística (INEC).

- Conectividad avanzada en zonas urbanas: La velocidad media de descarga para banda ancha fija alcanzó 157.29 Mbps (puesto 33 de 158 países). El uso de Internet llegó al 72.8% de la población en 2022.

- Infraestructura de datos existente: Colombia cuenta con 39 centros de datos y se han emitido lineamientos para el uso de servicios en la nube (Directiva Presidencial 03 de 2021).

Desafíos:

- Brecha de conectividad Urbano-Rural: Persiste la brecha entre el acceso a Internet en hogares de grandes áreas urbanas (72.20 %) y zonas rurales (32.22%).

- Necesidad de gobernanza de datos: Se requiere actualizar el Plan Nacional de Infraestructura de Datos (PNID) y el Modelo de Gobernanza de Infraestructura de Datos del MinTIC, y aumentar el presupuesto y financiación para invertir en computación en la nube y la infraestructura de

datos necesaria para el ecosistema digital.

El documento hace una anotación especial sobre la información falsa y el contenido sintético. La calidad de la información representa un problema relevante. El reporte cita que, durante la pandemia de COVID-19, el 76.68% de la información validada en internet fue falsa, y el 87.4% del contenido viral se difundió a través de redes sociales. Este fenómeno se agrava con la IA, y Colombia carece de un marco integral para la detección y remoción de contenidos infractores como discursos de odio o desinformación en Internet. La mitigación de la información falsa y el contenido engañoso fue señalada como una prioridad en las consultas públicas.

Recomendaciones estratégicas

Las recomendaciones de política se estructuran en dos áreas: gobernanza e institucionalidad (fortalecer el marco institucional y normativo) y creación de capacidades (promover la alfabetización digital y la formación especializada).

La tabla 1 resume dichas recomendaciones, enfocándose en las acciones de Alta y Media prioridad con horizontes de mediano y largo plazo, y aquellas de implementación continua, esenciales para el seguimiento estratégico por parte del gobierno nacional.

Tabla 1. Recomendaciones estratégicas para Colombia

Recomendación	Entidades Líderes	Prioridad	Marco Temporal	Objetivo estratégico y relevancia para el gobierno nacional
Gobernanza e Institucionalidad				
Consolidación de las instancias de gobernanza de la IA para la articulación inter-ramas (Ejecutivo, Legislativo, Judicial) y a nivel territorial.	Presidencia de la República, DNP, MinTIC, MinCiencias	**Alta**	Implementación continua	Fortalecer el rol de la Mesa Interinstitucional de IA para garantizar la coherencia y expedición de legislación que atienda a la política pública de IA y al marco constitucional. Incluir métricas de adopción ética en el FURAG (Formulario Único de Reporte de Avances de la Gestión).

Recomendación	Entidades Líderes	Prioridad	Marco Temporal	Objetivo estratégico y relevancia para el gobierno nacional
Fortalecimiento de infraestructura, gobernanza e interoperabilidad de datos públicos.	MinTIC, Cancillería y DANE	**Alta**		Promoción de la Digitalización y Gobernanza de Datos de Calidad para el Entrenamiento de Modelos de IA. Objetivo Estratégico: Fortalecer el ecosistema de datos, clasificado como "Diferenciador" (3.6/5), mediante la digitalización de conjuntos de datos críticos para el entrenamiento de modelos de IA y la garantía de su disponibilidad y calidad. Se debe fomentar la interoperabilidad y portabilidad de los datos y fortalecer la gobernanza, impulsando la creación de la figura del *Chief Data Officer* en más instituciones públicas. (PNUD, 2024).
Creación de espacios de diálogo sectorial y multiactor organizados por sectores estratégicos (agro, salud, etc.).	Presidencia de la República, DNP, MinTIC, MinCiencias, y Ministerios de cada sector	Largo plazo	Asegurar la inversión presupuestaria y privada (revisión de Decreto 1974 de 2019 sobre APPs - Asociaciones Público-Privadas) en infraestructura de cómputo y datos. Actualizar el PNID y Modelo de Gobernanza de Datos para adaptarse a las necesidades de IA, esencial para el desarrollo de sistemas confiables.	

Recomendación	Entidades Líderes	Prioridad	Marco Temporal	Objetivo estratégico y relevancia para el gobierno nacional
Fortalecimiento de capacidades de cooperación internacional del Gobierno Nacional en IA.	Presidencia de la República, Cancillería, MinTIC, MinCiencias e ICONTEC	Mediano plazo	Promover la adopción de IA en sectores clave y garantizar que el desarrollo tecnológico se alinee con los principios éticos y de derechos humanos de la UNESCO, mediante un enfoque *bottom-up* y colaboración multisectorial.	Asegurar la participación activa de Colombia en procesos globales de estandarización y gobernanza de IA, priorizando las necesidades nacionales y promoviendo el intercambio de mejores prácticas.
Revisión y actualización de la Ley 1581 de 2012 (Protección de Datos Personales).	Congreso de la República, MinCIT (Ministerio de Comercio, Industria y Turismo) y SIC (Superintendencia de Industria y Comercio)	**Alta**	Corto plazo	Modernizar el Régimen de Protección de Datos Personales para alinearlo con la rápida evolución de tecnologías como la IA, garantizando un marco legal eficaz que proteja los derechos personales y la diversidad cultural y territorial.
Creación de mecanismos exploratorios de regulación (sandbox) para IA.	MinTIC, MinCiencias, MinCIT y SIC	Media	Mediano plazo	Fomentar la innovación responsable permitiendo que desarrolladores prueben soluciones de IA en un entorno controlado y supervisado, acelerando la adopción tecnológica sin comprometer derechos fundamentales.
Creación de medidas e indicadores claves de desempeño (KPIs) para la	MinTIC, MinCiencias, DANE y DAFP (Departamento Administrativo de la	Medio	Corto plazo	Desarrollar indicadores estandarizados y verificables (auditorías, conocimiento de los usuarios, equidad percibida) para medir la transparencia,

Recomendación	Entidades Líderes	Prioridad	Marco Temporal	Objetivo estratégico y relevancia para el gobierno nacional
adopción ética de IA.	Función Pública)			equidad y responsabilidad en el uso de IA en la administración pública (vía FURAG).
Creación de Capacidades				
Aumento de la inversión en I+D+i (Investigación, Desarrollo e Innovación) especialmente en impactos sociales y ambientales de la IA.	MinCiencias, MinTIC, MHCP (Ministerio de Hacieda y Crédito Públic), MEN (Ministerio de Educación Nacional), Cancillería, ICONTEC	**Alta**	Implementación continua	Superar la debilidad estructural del bajo gasto en I+D+i (0.29% del PIB) mediante la destinación de partidas presupuestarias específicas, fondos concursables, e incentivos a empresas para el desarrollo de soluciones de IA con impacto social y ambiental positivo. Diseñar un Plan Nacional para Infraestructura Avanzada de IA, diferenciándola de la infraestructura de propósito general. El gobierno debe invertir en la creación y mejora de infraestructura crítica (centros de datos, supercomputadores de alto rendimiento) para fortalecer la soberanía digital (PNUD,2024).
				Ampliar y actualizar los programas del SENA y universidades, creando un sistema nacional de certificaciones que incluya un componente obligatorio de ética,

Recomendación	Entidades Líderes	Prioridad	Marco Temporal	Objetivo estratégico y relevancia para el gobierno nacional
Fortalecimiento de la oferta pública y privada de capacitación y certificación en IA.	MEN, MinTrabajo, MinTIC y SENA	**Alta**	Implementación continua	transparencia y sostenibilidad. Se debe alinear el Plan Nacional de Formación y Capacitación 2020-2030, integrando IA, ética pública y transformación digital en los programas de alta dirección y liderazgo. Asimismo, se requiere crear incentivos para la atracción de talento especializado y científicos de la diáspora colombiana. (PNUD, 2024)
Actualización de los currículos educativos para incluir capacidades técnicas y ética de IA.	MEN, MinTIC, MinTrabajo, MinCiencias, MinCultura, DNP	**Alta**	Corto plazo	Diseñar una política nacional de formación docente en IA e integrar contenidos éticos e interdisciplinarios (sesgos, privacidad) en la educación básica, media y superior, asegurando la pertinencia y el cumplimiento de la Ley General de Educación.
Fortalecer la formación y el acceso abierto a conocimiento sobre IA para toda la población.	MEN, MinCiencias, MinTrabajo, MinCultura	Medio	Mediano plazo	Crear un repositorio o biblioteca digital nacional de recursos educativos abiertos en IA (cursos, guías, juegos pedagógicos), con énfasis en la accesibilidad (idiomas, formatos para discapacidad) y poblaciones vulnerables.
				Articular la oferta educativa y la investigación en IA

Recomendación	Entidades Líderes	Prioridad	Marco Temporal	Objetivo estratégico y relevancia para el gobierno nacional
Creación de un repositorio centralizado de formación y proyectos de investigación sobre IA.	MinTIC, MinCiencias, MEN, SENA	Medio	Mediano plazo	del país (universidades, SENA, entidades públicas) en una plataforma única, facilitando el mapeo de capacidades y la identificación de brechas en investigación ética y sostenible.

Nota: Elaboración propia con apoyo de herramienta de inteligencia artificial generativa de Google.

Conclusiones

Colombia ha sentado bases sólidas para la adopción ética de la IA. Su hoja de ruta está claramente definida por las 106 acciones del CONPES 4144 de 2025 y las recomendaciones de política derivadas de este diagnóstico en profundidad realizado con la UNESCO, enfocadas en dos áreas: Gobernanza e Institucionalidad (consolidación de la Mesa Interinstitucional de IA, fortalecimiento de la infraestructura de datos y clasificación de riesgos sectoriales) y Creación de Capacidades (aumento de la inversión en I+D+i y actualización curricular con ética de IA). La visión es que la gobernanza ética y la innovación son plenamente compatibles y esenciales para asegurar un ecosistema tecnológico que promueva el bien público. Para materializar esta visión, el gobierno nacional debe priorizar la inversión continua y estratégica en I+D+i y en el talento humano con enfoque ético e inclusivo, transformando las debilidades estructurales identificadas en oportunidades reales de desarrollo sostenible e inclusivo para todos los ciudadanos.

Referencias

CEPAL. (2025). *Índice Latinoamericano de Inteligencia Artificial (ILIA) 2025*. Comisión Económica Para América Latina Y El Caribe. https://www.cepal.org/es/publicaciones/82514-indice-latinoamericano-inteligencia-artificial-ilia-2025

PNUD. (2024). *AILA: Evaluación del Panorama de la Inteligencia Artificial en Colombia*. UNDP. https://www.undp.org/es/colombia/publicaciones/aila-evaluacion-panorama-inteligencia-artificial-colombia

UNESCO (2025). *COLOMBIA: Metodología de evaluación del estadio de preparación*. https://unesdoc.unesco.org/ark:/48223/pf0000396015

***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–.*

***Andres R. Almanza J., M.Sc., CISM**. Chief Growth Officer en CISOS.CLUB, Investigador en Ciberseguridad SegInfo y Liderazgo. | Executive Certificate in Cybersecurity Leadership & Strategy by FIU University | Certificado como ISO 27001 Lead Implementer and 27005 Lead Manager from PECB | CISM, ITILv3, LPI| Certificado como Coach Profesional Internacional, Master in Leadership and Organizational Development with Coaching, Executive Master's in Leadership Skills Developed in Harvard, & Coach Profesional avalado por International Coach Federation| Profesional en Ingeniería de Sistemas | especialista en seguridad en redes y máster en seguridad de la información. Docente del programa de maestría de la Universidad Externado de Colombia y de la Universidad de las Américas en Ecuador. Creador de la Comunnidad CISOS.CLUB, CISOS-COLy CISOS-LATAM (Linkedin) y Miembro del comité editorial de la revista sistemas de ACIS.*

Los asuntos más relevantes de la IA en opinión de algunos expertos

DOI: 10.29236/sistemas.n177a5

Jeimy J. Cano M.

La primera pregunta dice, la inteligencia artificial ya no es una promesa. Es una realidad para las organizaciones. En este contexto, ¿cómo debe abordar este nuevo reto a las organizaciones? ¿Cuál sería el paso a paso para su implementación?

Guillermo Zegarra

Auditor Corporativo Seguros y Pensiones Credicorp Ltd.
Perú

Hoy la inteligencia artificial (IA) dejó de ser una promesa para convertirse en una realidad estratégica. Las organizaciones que entienden esto saben que la IA no es solo tecnología: es un habilitador que impulsa la transformación del modelo de negocio y la experiencia del cliente.

¿Cómo abordar este reto?

El primer paso es comprender que la IA debe estar alineada con la estrategia corporativa. Su propósito no es reemplazar personas, sino potenciar capacidades, mejorar procesos y colocar al cliente en el centro. Esto exige irradiar una cultura digital que permee toda la organización, traducida en acciones concretas y rutas de aprendizaje para cada rol.

Desde el punto de vista del talento, el desafío es doble: motivar y generar confianza. La IA será un facilitador que libera tiempo para actividades de mayor valor, pero requiere programas de upskilling y reskilling que preparen a los equipos para nuevos roles híbridos, como auditores digitales o analistas de datos.

¿Cuál es el paso a paso para su implementación?

- Definir la visión y gobierno
Establecer políticas corporativas sobre uso responsable, ética y protección de datos.
Crear un comité de IA que involucre TI, riesgos, auditoría y negocio.

- Construir cultura y capacidades
Comunicar que la IA es una ventaja competitiva.
Diseñar rutas de aprendizaje y fomentar la experimentación segura.

- Identificar casos de uso prioritarios
Quick wins con impacto medible: automatización de informes, análisis predictivo, atención al cliente.
Evaluar viabilidad técnica y regulatoria.

- Implementar pilotos y escalar
Probar en entornos controlados, medir resultados y ajustar.
Documentar aprendizajes para replicar.

- Monitorear y mejorar continuamente
Definir KPIs: eficiencia, reducción de errores, satisfacción del cliente.
Revisar riesgos emergentes como sesgos y ciberseguridad.

En nuestro caso, como parte de un conglomerado financiero, hemos adoptado políticas corporativas para garantizar un uso ético y seguro de la IA. En auditoría interna, estamos impulsando la transformación hacia auditores digitales, con iniciativas como:

- Generación automática de informes y observaciones.
- Construcción de matrices de riesgos y controles.
- Análisis masivo de datos para detección de anomalías.

Creemos en el aprendizaje haciendo, asumiendo errores y corrigiendo. La IA no sustituye el criterio profesional, lo potencia. El futuro no es esperar a que la tecnología llegue: es liderar su adopción responsable y estratégica.

Germán Noguera C.
CEO – Gerente ONC S.A.S.
Por un lado, para el abordaje de este reto, desde luego hay una dimensión técnica, y no será lo mismo la inteligencia artificial aplicada a una compañía de seguros, como comenta Guillermo, que aplicada a una organización del sector de la salud, a una empresa de ingeniería o a una entidad del gobierno. Lo cierto es que la inteligencia artificial hoy en día tiene que ver con todas las áreas.

Más allá de la dimensión técnica, es muy importante tener en cuenta la dimensión ética de los procesos (que es el tema que tratamos en la Comisión de ética de ACIEM), la gestión de riesgos y, desde luego, la dimensión humana.

La Comisión Europea (2019) insiste en la necesidad de una «IA confiable», construida sobre principios como la **transparencia, la equidad y el control humano continuo**.

El documento "Recomendación sobre la ética de la inteligencia artificial", adoptado en 23 de noviembre de 2021 por la UNESCO, reconoce (entre otros puntos) las repercusiones positivas y **negativas** profundas y dinámicas de la inteligencia artificial (IA) en las sociedades.

Y propone unos principios y unos valores que deben ser considerados en todas las etapas del ciclo de vida de los sistemas de IA.

Los principios propuestos son:

- Proporcionalidad e inocuidad
- Seguridad y protección
- Equidad y no discriminación
- Sostenibilidad
- Derecho a la intimidad y protección de datos
- Supervisión y decisión humanas
- Transparencia y explicabilidad
- Sensibilización y educación
- Gobernanza y colaboración adaptativas y de múltiples partes interesadas
- Responsabilidad y rendición de cuentas

También quisiera mencionar la existencia de la norma ISO 42001: 2023 Sistemas de Gestión de Inteligencia Artificial que puede ser una buena referencia, con aportes para asumir el reto.

En cuanto a la dimensión humana de las organizaciones, es fundamental una adecuada gestión del cambio (en las empresas ya se tiene experiencia en procesos de cambio con la adopción del aspecto ambiental en la gestión empresarial, o los temas de transparencia, anticorrupción, o la responsabilidad social). Sin embargo, una diferencia en relación con la IA es la velocidad de evolución de las tecnologías que podría llevarnos de una gestión del cambio puntual o por proyectos a un esquema de gestión continua del cambio, que implica mayores velocidades de reacción y capacidad de adaptación.

También, frente a la adopción de la IA, es necesario tener en cuenta que, a diferencia de las empresas de hace 50 años, hoy en día en la gestión empresarial es necesario tener en cuenta, además de los accionistas y el resultado financiero, a todas las partes interesadas, como los empleados, los proveedores, clientes, el medio ambiente, autoridades, entre otros. Y los temas de responsabilidad social y ESG se deben considerar cada vez con mayor relevancia.

Guillermo Zegarra
Nosotros tratamos de entender el momento en el que estábamos a la hora de adoptar la inteligencia artificial. Y nos preguntamos y tratamos de resolver para qué, para quiénes y a través de quiénes la vamos a utilizar. Entonces entramos dos pilares muy sencillos, el primero entender nuestra gente, en qué momento están. Nos pudimos dar cuenta lo que las personas entendían por inteligencia artificial, qué estaban haciendo con la inteligencia artificial antes de estar en la empresa. Y era más famoso Samsung y los proveedores de dispositivos móviles, los cuales lograron transmitirle a las personas su mensaje de inteligencia artificial.

Jeimy J. Cano M.
¿Qué decirles a las personas que ahora se enfrentan y compiten frente a una tecnología que aprende más rápido que ellas?

Fabio Alexander Rojas Roldán
Vicepresidente de Auditoría de TI en KPMG
Es una pregunta difícil de responder, en la medida que toca ciertas fibras sensibles. En mi opinión debemos iniciar por sincerar el discurso; la inteligencia artificial está transformando profundamente el mercado laboral y reemplazará empleos tradicionales como parte de una evolución normal. La sociedad debe adaptarse a esta tecnología creando nuevas oportunidades laborales y trabajando en conjunto: empresas, gobierno, academia y ciudadanía. No basta con responsabilizar solo a las empresas para que creen nuevas plazas de trabajo, ni con recomendar a las personas hacer cursos de inteligencia artificial; es necesario un enfoque integral, sincero, sin alarmismo, pero célere; para gestionar este cambio. Debemos generar espacios de diálogo que involucren a todos los actores y enfrentar la disrupción tecnológica de manera realista.

Te doy el ejemplo de una compañía que implementa RPA (Automatización de procesos basada en robótica), con esto logra identificar un centenar de tareas rutinarias que hace un numero X de personas a diario y que son tareas que pueden hacer de inicio a fin estos “robots”, los cuales tienen un muy bajo nivel de error, trabajan sin parar 7*24, no tienen gastos adicionales de nómina y generan incrementos en productividad que son imposibles de lograr con la fuerza laboral basada

en el humano; sus ventajas son muy grandes en comparación y los beneficios para la empresa traen todo el sentido necesario para que esto sea implementado. No es posible que la organización genere la misma cantidad de nuevas tareas y puestos de trabajo, que los que son reemplazados por los Bots y no debería ser responsabilidad única de la empresa, los trabajadores deberían poder encontrar nuevas actividades que la sociedad en su conjunto crea para responder a esta nueva realidad y para eso se necesita un programa articulado entre el Gobierno, la Academia y la Industria.

Guillermo Zegarra

La inteligencia artificial no debe verse como un sustituto, sino como un complemento estratégico. Las organizaciones y las personas estamos lo suficientemente maduras para abordar este reto con transparencia y responsabilidad. La IA transformará roles, automatizará tareas y reducirá cargas operativas, lo que se traduce en mayor eficiencia. Pero esto no significa que el talento humano pierda relevancia; al contrario, se potencia.

¿Cómo lo estamos haciendo?

En nuestro contexto, hemos incorporado agentes de IA para optimizar procesos operacionales, y los beneficios son claros: más agilidad, mejor calidad y capacidad de análisis. Esto se acompaña de tres pilares fundamentales:

- Comunicación clara y transparente
  Explicar el propósito de la IA, sus beneficios y límites, para reducir temores y generar confianza.

- Capacitación y reskilling
  Preparar a las personas para nuevas competencias: pensamiento crítico, supervisión de IA, ética y creatividad. Según el World Economic Forum, el 50% de los trabajadores necesitará algún tipo de reskilling antes de 2027 debido a la adopción de tecnologías emergentes.

- Políticas y gobierno de IA
  Definir marcos éticos, gestión de riesgos y mecanismos de supervisión para garantizar un uso responsable - marcos claros de transparencia y protección de datos.

Además, es clave generar espacios de escucha activa, donde los colaboradores puedan expresar dudas y temores, y responder con planes concretos. La IA libera tiempo para tareas de mayor valor, mejora la calidad del trabajo y la experiencia del cliente. El verdadero desafío no es competir con la IA, sino aprender a trabajar con ella.

Considero que la IA no reemplaza el criterio humano, lo amplifica. Las organizaciones que adopten esta visión, combinando tecnología con talento, serán las que lideren el futuro.

Jeimy J. Cano M.
Guillermo, sólo como para contrastar, con esta respuesta de Fabio, ¿la nómina de ustedes sí se va a reducir o no?

Guillermo Zegarra
Hasta el momento, no ha sucedido eso. Hemos obtenido recursos con el compromiso de explorar y desarrollar casos de uso de IA.

Germán Noguera C.
Sí, estoy de acuerdo con lo que dice Fabio, es necesario reconocer y aceptar esta realidad y ver como sociedad cómo vamos a reaccionar ante esto, para minimizar los impactos en las personas.

Si yo soy una organización y no me actualizo, no utilizo las herramientas y mi competencia sí, pues seguramente voy a salir del mercado y el impacto va a ser todavía mayor porque entonces voy a tener a todos mis empleados desempleados porque desaparecerían esas fuentes de trabajo.

Entonces es una realidad que las empresas enfrentan. Como en el caso del granjero, si la empresa no adopta las herramientas de tecnología y su vecino sí, al final saldrá del mercado porque el vecino será capaz de vender a menor precio. El mercado y la competencia nos llevan a adoptar estas herramientas y debemos ser realistas y contarle a la gente qué está pasando, para dónde va la empresa y para dónde va el negocio, para que todos estemos enterados de lo qué está pasando y que, si no se avanza, la empresa va a desaparecer con los puestos de todos.

Fabio Alexander Rojas R.

Estoy plenamente alineado con lo expuesto por Germán, pero consi-

dero pertinente ilustrar mi perspectiva mediante el ejemplo de un agricultor. Circula en internet un video que muestra a un granjero que ha implementado una máquina operada mediante inteligencia artificial. En dicha granja no interviene mano de obra humana: la IA se encarga de alimentar al ganado, calcula con exactitud la ración necesaria y la distribuye automáticamente. El agricultor gestiona todo desde una aplicación móvil, recibiendo información precisa y constante sobre el consumo de alimento, los períodos de descanso, posibles alertas sanitarias, entre otros indicadores. Incluso, el robot realiza el ordeño regular de las vacas, cubriendo tareas repetitivas que suelen ser poco atractivas para los trabajadores y que, de realizarse manualmente, implicarían costos significativamente mayores a largo plazo. El propio granjero manifiesta disponer de mayor tiempo libre y haber incrementado sustancialmente sus beneficios. Resultaría poco razonable exigirle que no adopte esta tecnología para evitar el desplazamiento laboral o que genere nuevos puestos innecesarios bajo el nuevo modelo operativo, ya que ello solo ocasionaría incrementos de costos sin justificación. Además, este sistema proporciona mejoras en la calidad y costos de los productos, beneficiando así a los consumidores.

No obstante, la cuestión central radica en que, en este caso, es únicamente el agricultor quien expresa su opinión; no se observa una intervención del Ministerio de Agricultura abordando cómo el gobierno reconoce esta situación e implementa políticas para que los productores agrícolas puedan desarrollar nuevas competencias, incentivando la creación de industrias o modelos productivos innovadores que generen empleos, ni tampoco la existencia de programas claros que integren a la academia como apoyo en este proceso.

Jeimy J. Cano M.
¿Cómo prepararse para enfrentar este nuevo riesgo? ¿Se requieren nuevas capacidades y habilidades? ¿Por dónde empezar?

Fabio Alexander Rojas R.
Para gestionar este tipo de temas, es fundamental contar con un marco de gobernanza claro. Lo primero es reconocer que existe un riesgo, comprenderlo y desarrollar una estrategia para enfrentarlo.

Esto implica establecer políticas de transparencia respecto al uso de algoritmos —políticas que deben estar diseñadas desde el inicio como parte del gobierno corporativo, no como algo opcional— reconociendo la importancia de la transparencia y el enfoque ético en estos casos.

Además, es esencial trabajar con las personas. Existen muchos casos en los cuales humanos e inteligencia artificial trabajan juntos y los humanos mantienen un rol central,

por lo que debemos aprender a utilizarla, enfrentar sus retos y aprovechar sus beneficios. En nuestra organización y en otras hemos notado que la IA puede resultar muy persuasiva, expresando resultados convincentes incluso si son erróneos. Si las personas no comprenden cómo funciona ni pueden ejercer un juicio crítico sobre estas tecnologías, el riesgo se intensifica, porque resulta fácil dejarse convencer por la IA, incluso cuando se equivoca.

Por eso, el marco de gobierno debe orientarnos a comprender los riesgos, conocer cómo se desarrollan y operan estas tecnologías, y garantizar su uso transparente y auditable. Debemos preparar a las personas para interactuar con nuevas generaciones de algoritmos que presentan capacidades distintas a las conocidas, ayudándolas a adaptarse a esta nueva era tecnológica y ofreciéndoles herramientas para orientar su juicio y aprovechamiento responsable.

Mientras sigamos siendo responsables por el manejo de estas tecnologías, tenemos que seguir fortaleciendo nuestras habilidades.

Aunque hablamos de inteligencia artificial, el desarrollo personal sigue siendo clave para mantener bajo control los riesgos asociados, hasta donde sea posible. En el futuro, cuando llegue la computación cuántica, seguramente tendremos que abordar nuevos desafíos.

Guillermo Zegarra

La inteligencia artificial (IA) no solo trae oportunidades, también introduce riesgos que deben ser gestionados con visión estratégica. El primer paso es **reconocer el riesgo y mapearlo** dentro de la estrategia corporativa.

Hoy, este no es un tema menor: según el último informe *Risk in Focus* del Instituto Global de Auditores Internos, los dos principales riesgos identificados por las organizaciones son **ciberseguridad** y **disrupción digital**, donde la IA ocupa un lugar central.

¿Qué implica esto para las empresas?

Significa que debemos incorporar la IA en tres pilares:

- Gobierno corporativo: definir políticas claras sobre uso responsable, ética y transparencia.
- Gestión de riesgos: evaluar impactos potenciales, sesgos, errores y riesgos regulatorios.
- Control interno y auditoría: desarrollar capacidades para auditar la IA y garantizar que sus decisiones sean confiables.

En nuestro caso, hemos implementado casos de uso en auditoría que nos han permitido mayor alcance y cobertura, liberando tareas repetitivas y enfocándonos en análisis de valor. Sin embargo, no todos los casos funcionan; algunos se desestiman tras pruebas piloto,

lo que confirma que la IA requiere experimentación y evolución continua.

¿Se necesitan nuevas capacidades?

Definitivamente. Las organizaciones y las personas deben invertir en reskilling y upskilling. Esto incluye, por ejemplo, conocimiento en marcos de referencia como NIST AI Risk Management Framework útil tanto para implementar como para auditar IA.

¿Por dónde empezar?

1. **Reconocer el riesgo en la estrategia corporativa**.

2. **Definir políticas y gobierno de IA**.

3. **Capacitar y preparar al talento** para roles híbridos.

4. **Implementar pilotos controlados** y medir resultados.

5. **Adoptar marcos internacionales** para gestión y auditoría de IA.

La IA no sustituye el criterio humano, lo amplifica. Las organizaciones que combinen tecnología con talento, gobernanza y ética serán las que lideren el futuro.

Germán Noguera C.
En la Comisión de Ética de ACIEM hemos reflexionado sobre este tema de los riesgos en la IA, y la ética como elemento que contribuye a mitigar los riesgos, y pensamos que el tema debe abordarse desde tres niveles: La ética de los desarrolladores, la ética de los promotores o vendedores de la tecnología y la ética de los usuarios.

Cuando hablamos de los riesgos hay también una diversidad de escenarios, se habla de los sesgos, de las alucinaciones, de la fuga de la información, o de la pérdida de confidencialidad de la información, entre otros riesgos. Las herramientas de IA pueden generar, por accidente o por error, información que es falsa pero que es convincente, pero también puede haber una mala intención en los mismos desarrolladores o personas que utilizan la IA para, por ejemplo, generar noticias falsas, rumores o para dañar la reputación de alguien.

Tenemos hoy en día empresas dedicadas al ciberdelito. Ya no se trata de un hacker actuando de manera individual, sino que nos enfrentamos a organizaciones completas con equipos de trabajo, igual que una empresa, pero dedicadas a hacer las cosas malas.

Debe haber un esfuerzo de la sociedad por generar una conciencia universal. Universal en el sentido de llegar a todos, pero con el propósito de crear en cada persona una conciencia individual en cuanto a los riesgos asociados a la inteligencia artificial.

Entender que existen riesgos en diferentes ámbitos, en la casa y en las familias, en las organizaciones o en los gobiernos. Todas las personas deberían ser conscientes de los riesgos que existen asociados a la IA.

También se requiere un esfuerzo desde el punto de vista político y de gobernanza, porque no nos podemos quedar con confiar en que la gente es buena y es ética y que no van a actuar mal. Es necesario ir más allá, y esto tiene que ver con temas de legislación, de vigilancia y control y de sanción.

Se debe buscar que lo no ético resulte costoso (desde el punto de vista de sanciones). Hoy vemos que hay un problema asociado a esto y es que el ser no ético no es costoso. Si cuando cometo el fraude o hago la trampa o incurro en un acto corrupto, al final nadie me sanciona o la sanción es mínima, se están promoviendo estos comportamientos. Pero si el costo del mal comportamiento es alto, se reduce la probabilidad de que haya personas actuando de manera incorrecta.

Se debe buscar que las sanciones por el mal uso de la IA sean suficientemente fuertes, que exista un sistema robusto para poder establecer las responsabilidades de las personas detrás de la inteligencia artificial, porque entonces no se podrá argumentar que lo ocurrido fue culpa del programa, o fue culpa de la inteligencia artificial o de la herramienta, para librarse de la responsabilidad. Con un sistema robusto, que permita establecer las responsabilidades y las sanciones de modo que lo no ético sea costoso para las empresas o personas infractoras. Es similar a lo que ha pasado con lo ambiental: si contaminar no cuesta, contamino; pero cuando se establecen mecanismos para cuantificar el daño ambiental, identificar al responsable y hay sanciones y multas, entonces se frena un poco el tema.

Pero también, se debe pensar en mecanismos que promuevan o premien el buen comportamiento. Hoy lo vemos cuando hablamos de transparencia, y Guillermo quien es contador, conoce, por ejemplo, las NIIF, las Normas Internacionales de Información Financiera, que establecen un marco de referencia para que todas las empresas reporten la información financiera y las notas de revelación, de una manera que sea clara para todos. Ahí estamos aportando un elemento de transparencia. O cuando hablamos de los sistemas de lucha contra el lavado de activos y o la financiación del terrorismo, para los que hay unos reportes y unas certificaciones, que todos conocen y pueden verificar.

Con el tema de la inteligencia artificial, tendremos que llegar a algún modelo de auditoría y de certificación por terceros, más allá de la propia organización haciendo sus

declaraciones, para que todas las partes interesadas puedan ver que ese desarrollador, ese promotor de la herramienta o ese usuario de la herramienta están haciendo uso adecuado de todos estos elementos de la inteligencia artificial.

En resumen, se debe trabajar en la construcción de una cultura profesional ética, pero ir más allá con las acciones para detección de prácticas indebidas y mecanismos para identificar y sancionar a los responsables, y también con los mecanismos para que, con declaraciones voluntarias y sistemas de certificación por parte de terceros, se informe a los usuarios de la IA, en relación con los riesgos y la correcta utilización de la IA.

Jeimy J. Cano M.
La cuarta pregunta que hace énfasis particularmente en uno de esos riesgos: "las alucionaciones", uno de los riesgos más mencionados en la literatura, y al que más expuestos están tanto las personas como las organizaciones frente a la IA Generativa. Así las cosas, la pregunta es, ¿cómo enfrentar el reto de las alucinaciones? ¿Cómo hacer más confiable las respuestas de la IA Generativa?

Germán Noguera C.
Las alucinaciones tienen que ver con el riesgo de generación de información falsa o equivocada, pero creíble. Y para usar el mismo término, pero con otro significado, podemos decir que alucinante es también la velocidad con la que se avanza en esta ruta del desarrollo de la IA Generativa.

En la actualidad, la IA generativa es algo que puede apoyar nuestras labores, de manera complementaria, como una especie de copiloto, que consultamos o utilizamos para hacer nuestras tareas. Pronto tendremos el gemelo digital, con mayores capacidades y autonomía, que no sólo ayuda, sino que también ejecuta, hasta llegar a agentes autónomos. Y en la medida en que las herramientas se alejan de la participación humana y tienen mayor autonomía, las posibilidades de control o verificación se reducen y el riesgo de alucinaciones se incrementa.

Seguramente, hay soluciones técnicas que serán provistas por parte de las mismas herramientas, como controles cruzados o verificaciones múltiples. Sin embargo, más allá del aspecto técnico, creo que es importante incentivar a quienes desarrollan IA para que prioricen la ética por encima del lucro y la presión de resultados.

Hemos visto en otros negocios, como en la industria automotriz o en la aeronáutica, que por el afán de lanzar el producto antes que los competidores no se hacen todas las verificaciones necesarias, y luego, cuando ocurren accidentes o se detectan problemas de operación, se corre a hacer reparaciones, ajustes o cambios para los compradores.

Esto también puede ocurrir en la industria de desarrollo de la IA Generativa, con la aparición de productos en el mercado que no han sido probados suficientemente. Por ello, retomo lo dicho en alguna respuesta anterior, en cuanto a la importancia de hacer que lo no ético sea costoso, y que las empresas (al nivel del desarrollador de las herramientas, del que las vende o la promueve y también del que las usa) asuman sus responsabilidades, y se apliquen las sanciones y castigos; y, de manera paralela, crear incentivos que hagan que lo ético sea rentable y que lo riesgoso sea costoso, porque las empresas reaccionan antes estos mecanismos.

En la Comisión de ética de ACIEM hablamos, en el caso de la ingeniería, de incorporar la ética en la formación de los ingenieros, que la ética forme parte, no solamente del programa de formación como una materia, sino que esté involucrada en las diferentes materias de la carrera, para que en el ejercicio profesional la ética sea un elemento cotidiano y transversal en todas las actuaciones del ingeniero.

Y entonces, pienso que la ética también debe estar formando parte de esos elementos transversales y cotidianos del negocio de la inteligencia artificial, en todas las partes que intervienen.

Lo anterior, complementado con el aspecto de la transparencia pública como parte del modelo de negocio, que va de la mano de esa conciencia sobre los riesgos que debería existir en todos los usuarios (también mencionada en respuesta anterior).

Me gusta hacer el paralelo con otros temas organizacionales. Por ejemplo, en lo ambiental, cuando una empresa tiene el sallo ambiental, los clientes tienen preferencia por comprar sus productos; y cuando hay una noticia de que esa empresa contamina, el mercado mismo se encarga de sacarlo.

Y en relación con este tema de enfrentar el reto de las alucinaciones, creo que debería haber algo parecido, de modo que cuando la empresa sea transparente, y diga la verdad sobre su herramienta, lo que está haciendo, las limitaciones y grado de confiabilidad, o el uso que está dando a la herramienta, le vaya mejor que cuando no lo hace, o cuando oculta la información.

Dicho lo anterior, creo que los temas de transparencia y ética son elementos fundamentales para enfrentar el reto de la alucinación.

Jeimy J. Cano M.
¿Cómo han visto ustedes en su implementación el tema de las alucinaciones? ¿Cómo han enfrentado este reto?

Guillermo Zegarra
Las alucinaciones son uno de los riesgos más críticos en el uso de modelos de IA, especialmente en

entornos regulados como el nuestro. Las hemos experimentado en fases iniciales, principalmente cuando trabajamos con modelos de *Machine Learning* en procesos acelerados, con datos insuficientes o poco depurados.

Esto generó resultados inconsistentes y confusos, lo que nos obligó a retroceder, corregir y reforzar nuestras prácticas.

Lo abordamos partiendo de un principio fundamental: reconocer el riesgo y actuar con ética y transparencia. Esto implica asumir que las alucinaciones pueden ocurrir y establecer mecanismos para prevenirlas y corregirlas. Además, trabajar en un entorno corporativo de colaboración ayuda a aprender. Trabajamos en tres frentes:

Gobierno y calidad de datos: La base para reducir alucinaciones es contar con datos confiables. Implementamos prácticas de gobierno de datos apoyadas en marcos como DAMA-DMBOK, que nos permiten gestionar calidad, obsolescencia y trazabilidad. Sin datos limpios, no hay IA confiable.

Parametrización y validación continua: Ajustamos factores de configuración y realizamos pruebas iterativas para detectar desviaciones. Involucramos perfiles diversos —técnicos, de negocio y auditoría— para democratizar la revisión y enriquecer la interpretación de resultados.

Ética y gobernanza: Establecimos principios claros para reconocer errores y retroceder cuando sea necesario. Esto incluye protocolos para comunicar hallazgos y corregir modelos antes de su escalamiento.

¿Por qué es crítico? Las alucinaciones no solo afectan la calidad del análisis, sino que pueden comprometer decisiones estratégicas, la relación con clientes y el cumplimiento regulatorio. En nuestro sector, un error en la información enviada al regulador puede tener consecuencias significativas.

Este enfoque nos permite avanzar con seguridad, reconociendo que la IA es poderosa, pero requiere control, supervisión y mejora continua.

Por ello creo que las alucinaciones son un riesgo real, pero gestionable. La clave está en combinar gobierno de datos, validación técnica y principios éticos, asegurando que la IA sea un aliado confiable y no una fuente de vulnerabilidad.

Como dijo Germán, creo, y también Alex en su momento, esto avanza muy rápido. Entonces, en verdad, no puedes detenerte mucho tiempo porque también hay otros que van avanzando en tu mismo sector. Entonces tienes que avanzar y permitirte... Time to market, time to market. Sí, eso. Y permitirte equivocarte, pero corregir rápido y lo principal, identificarlo y salir e ir mejo-

rando. Creo que, como lo dije al principio, en la primera pregunta, o sea, Yo, por lo que he visto, por lo que hemos hecho nosotros, por lo que hemos experimentado, creo que hay que dar el paso y seguramente que al principio nos equivocaremos algo, pero después seguro que agarraremos velocidad en la ejecución y una vez que agarremos velocidad en la ejecución, pues seguramente que podremos hacer muchos, tendremos muchos casos de uso y experimentos que nos ayuden a mejorar.

Fabio Alexander Rojas R.
Nuestros resultados parten de una estrategia que primero reconoce el riesgo real de alucinación en la IA. Este riesgo no nos toma por sorpresa; somos plenamente conscientes de su existencia y actuamos en consecuencia. Basándonos en nuestra experiencia, consideramos fundamental fortalecer el criterio de las personas usuarias: deben entender que la IA puede fallar y evitar confiar plena e indiscriminadamente en sus respuestas.

Las personas expertas pueden evaluar mejor los resultados ofrecidos por la IA, pues su criterio especializado resulta esencial. Así, el enfoque no es universal, sino adaptado a los conocimientos y competencias requeridos tanto por quienes usan la herramienta como por la propia tecnología.

Desde el punto de vista técnico, restringir el dominio de aplicación y entrenamiento ha sido clave. Por ejemplo, entrenamos nuestra inteligencia artificial siguiendo metodologías propias y específicas: si el servicio es de auditoría, enfocamos el entrenamiento en ese campo; si es consultoría, lo acotamos a dicha área.

Cuando la IA debe abarcar te-mas demasiado amplios, aumenta el riesgo de alucinar, ya que intenta ofrecer respuestas, aunque no cuente con suficiente información precisa. Limitar el alcance temático ayuda a reducir este comportamiento erróneo.

La calidad de los datos también juega un papel crucial. Proveer información confiable facilita que la IA priorice respuestas correctas y minimiza la necesidad de "inventar" cuando encuentra vacíos informativos. Además, invertir en algoritmos bien diseñados y en técnicas adecuadas fortalece el entrenamiento y robustez de la IA en los ámbitos deseados.

En definitiva, combinar estos factores reduce el riesgo de alucinación. El verdadero problema no es tanto que la IA alucine, sino que se confíe ciegamente en esos resultados y se tomen decisiones basadas en ellos sin cuestionarlos.

Nuestra experiencia demuestra que esta estrategia permite gestionar mejor el fenómeno de la alucinación y disminuye significativamente su impacto.

Jeimy J. Cano M.
Bien interesantes los diferentes puntos de vista y las reflexiones que hemos tenido. Vamos para la última pregunta de nuestra sesión de hoy. Por tanto, ¿cómo visualizan el futuro a 2030 de la IA generativa? ¿Nuevos desarrollos? ¿Qué temas ven? ¿Qué podemos esperar y qué cosas nos pueden sorprender? ¿Qué han visto ustedes y qué están viendo hacia adelante?

Guillermo Zegarra
La inteligencia artificial generativa ya no es una promesa: es una realidad en expansión que está transformando industrias, modelos de negocio y la forma en que interactuamos con la tecnología. Sin embargo, lo que viene hacia 2030 será aún más disruptivo. Creo que vendrán:

IA generativa como sistema nervioso digital, para 2030, la IA no será solo una herramienta, sino el sistema nervioso digital del mundo humano. Los modelos monolíticos darán paso a arquitecturas modulares —"grafos de especialistas"— que permitirán precisión, eficiencia y seguridad a gran escala. Además, veremos agentes personales con memoria vital cifrada, capaces de recordar no solo lo que dijimos, sino el contexto emocional y la intención detrás de cada interacción.

Interacción humana – IA hiperpersonalizada, la IA comprenderá tono de voz, expresión facial y postura corporal, anticipando necesidades antes de que las verbalicemos. Esto abrirá paso a gemelos contextuales y experiencias inmersivas 4D en educación, salud y entretenimiento, casi indistinguibles de la realidad física.
Impacto económico y adopción masiva, Según PwC y McKinsey, la IA podría aportar entre 13 y 15 billones de dólares al PIB global para 2030, equivalente a un 14% del total. Solo la IA generativa se estima que inyectará 19.9 billones de dólares, representando el 3.5% del PIB mundial, con un crecimiento anual del 46%.

Computación cuántica: el acelerador definitivo, lo que realmente puede sorprendernos es la convergencia entre IA y computación cuántica. Esta tecnología, que parecía lejana, se proyecta como operativa en los próximos cinco años. Su capacidad para procesar múltiples combinaciones simultáneamente permitirá entrenar modelos complejos en horas, no semanas, y resolver problemas de optimización imposibles para la computación clásica. Esto impactará áreas críticas como simulación molecular, ciberseguridad y desarrollo de nuevos materiales.

Riesgos y gobierno, el avance exponencial traerá desafíos: sesgos, alucinaciones, uso malintencionado (deepfakes) y riesgos regulatorios. Por ello, el gobierno y la ética seguirán siendo pilares estratégicos para garantizar un uso seguro y responsable.

Jeimy J. Cano M.

En mi ejercicio de pronósticos de seguridad/ciberseguridad[1] se advierte una tendencia denominada "crimen sin rostro" es decir, el uso de la inteligencia artificial generativa con agentes autónomos, para materializar acciones criminales, con sólo ejecutar un prompt y listo. Esa es una realidad que no nos puede tomar por sorpresa.

Guillermo Zegarra

Sí, ahí estoy totalmente de acuerdo contigo. Eso no nos puede tomar por sorpresa. Son temas que creo ya tenemos evidencia, y no podemos quedarnos sin reacción y obviamente no tener la resiliencia suficiente como para salir adelante en estos casos. Pero definitivamente no podemos esperar. Son temas que ya los conocemos.

Germán Noguera C.

El 2030 parece lejos, pero si hacemos cuentas, ya en unas semanas llega el primero de enero de 2026, y estaremos a cuatro años de 2030. En todo caso, creo que en esos cuatro años vamos a ver unos cambios muy importantes, teniendo en cuenta la velocidad a la que avanza el desarrollo de estas tecnologías.

Hace unas semanas estaba preparando un artículo que fue publicado en la revista ACIEM sobre ética e inteligencia artificial, y encontré unas definiciones que seguramente para los ingenieros de sistemas no son desconocidas, pero para mí fueron algo nuevo, y se hablaba de la evolución de la inteligencia artificial, desde la IA, como la conocemos hoy, hacia la Inteligencia Artificial General (AGI), que tendrá unas capacidades similares a las de los humanos, con posibilidad de realizar tareas intelectuales al nivel en que un humano las realiza, con capacidad de aprender, razonar y adaptarse a nuevas situaciones. Indicaba el artículo que actualmente no existe una verdadera AGI, pero que se llevan a cabo esfuerzos de investigación y desarrollo.

Muy posiblemente en cuatro o cinco años, y apalancados en la computación cuántica, ya estaremos en

1 Cano, J. (2025). Pronósticos de seguridad/ciberseguridad 2026. Entre señales débiles y sorpresas predecibles. Blog IT-Insecurity. https://insecurityit.blogspot.com/2025/10/pronosticos-de-seguridadciberseguridad.html

ese mundo y comenzando algo que el articulo denominaba como Super Inteligencia Artificial (ASI), que sería una tecnología que supera la capacidad humana, con posibilidad de resolver problemas que actualmente están más allá de las capacidades de los humanos. Esto es un nivel de evolución de la IA teórico, en gran medida objeto de debate y especulación. No sé si en dos, tres, cinco o diez años estemos allá, pero es hacia dónde vamos y es lo que viene.

Fabio Alexander Rojas R.
Tengo principalmente dos perspectivas. Por un lado, una visión optimista y llena de esperanza respecto a nuestro trabajo y el esfuerzo por ayudar a las organizaciones en ese proceso. Imagino que hacia 2030 la inteligencia artificial estará integrada completamente en nuestra rutina diaria, tanto en lo personal como en el ámbito organizacional, con empresas habituadas y colaborando estrechamente con esta tecnología. Preveo organizaciones cada vez más dependientes de la IA, pero también mucho más eficientes, logrando mejores resultados y adquiriendo mayor protagonismo en los contextos económico y social. Considero que la economía será liderada por la inteligencia artificial, que se convertirá en el motor principal que determine las reglas del juego mundial. Las personas estarán cada vez más familiarizadas e interesadas en el tema, y la sociedad girará en torno a la IA, generando productos y servicios más cercanos a los clientes y al mercado, perfeccionando las dinámicas sociales, empresariales e incluso gubernamentales.

Sin embargo, si no actuamos, existe otra perspectiva que realmente me preocupa y aquí expreso mi opinión: la IA es una herramienta muy poderosa para una especie que aún no está preparada. Creo que quienes tienen intenciones negativas llevan cierta ventaja, ya que, mientras nosotros debatimos consensos y trabajamos en asuntos complejos como la ética, normas, leyes y procedimientos —lo cual toma tiempo—, ellos no se ven limitados por estas consideraciones y aprovecharan las ventajas de tecnologías avanzadas como la IA (combinada con computación cuántica). Estas herramientas parecen estar a favor de quienes buscan hacer daño. Si no comenzamos a prepararnos y tomar decisiones desde ahora, es posible que en 2030 vivamos en una sociedad rezagada frente al mal uso de la inteligencia artificial, pues el avance de quienes actúan sin restricciones es mucho más rápido.

Estoy convencido de que quienes estamos aquí representamos la responsabilidad de evitar estos riesgos y debemos tomarlos muy en serio. En cuatro años podemos perder una gran ventaja, ya que el avance en IA en ese tiempo equivale a años luz. No debemos permitir que nos superen, porque las ventajas de eficiencia, resultados y

costos que ofrece la inteligencia artificial son enormes y hacen que el trabajo sea más interesante al liberarnos de tareas rutinarias. Sin embargo, los riesgos asociados evolucionan tan rápidamente que, en comparación con otros hitos históricos, considero que 2030 es una frontera donde debemos actuar con seriedad desde hoy.

Introducción a los riesgos algorítmicos

Repensando los modelos de seguridad y control en la era de la inteligencia artificial

DOI: 10.29236/sistemas.n177a6

Resumen

La acelerada incorporación de la inteligencia artificial (IA) y los agentes autónomos obliga a las organizaciones a revisar detalladamente los resultados de su implementación, ya que los modelos basados en *machine learning* generan respuestas que muchas veces no corresponden a una lógica esperada, sino a un patrón estadístico predeterminado. Esta nueva realidad ha dado lugar al riesgo algorítmico, un riesgo emergente definido como la posibilidad de daño, pérdida financiera o afectación de la reputación empresarial que surge del uso, despliegue o explotación maliciosa de sistemas de IA, que hace evidente las limitaciones propias de los modelos tradicionales de seguridad y control. Por tanto, este artículo propone adoptar una perspectiva ampliada de seguridad que incluye la autenticidad, la utilidad y la posesión como nuevos elementos a revisar, además de introducir el concepto de auditoría algorítmica como un proceso de aseguramiento crítico previo al despliegue de iniciativas con inteligencia artificial. Finalmente, este documento presenta un enfoque holístico de la seguridad denominado “Confianza por Diseño”, que opera como marco integrador que vincula la ética, la equidad y la explicabilidad, junto con la seguridad y la privacidad, como una postura proactiva para la gestión de riesgos empresariales en la era de la IA.

Palabras clave

Riesgo algorítmico, Confianza por diseño, Agentes autónomos, Integridad semántica, auditoría algorítmica

Introducción

La inteligencia artificial avanza de forma acelerada, y su incorporación en prácticamente todos los aspectos de la dinámica social, hace que tanto personas como organizaciones comiencen a revisar con detalle sus resultados y los efectos de su implementación, comoquiera que las respuestas de estos modelos (algoritmos basados en *machine learning*) muchas veces no corresponden ni a un contexto, ni a una lógica concreta y esperada, sino a un patrón estadístico previamente establecido en la programación inicial de la iniciativa (Janapa Reddi, 2025).

En este sentido, las iniciativas basadas en inteligencia artificial que se incorporan en la actualidad, no son capaces de distinguir entre datos contaminados o maliciosos y comandos legítimos, lo que no les permite interpretar la información de manera correcta, válida y conforme a la realidad o al contexto para el cual fueron diseñadas, creando una brecha de integridad semántica que genera necesariamente alucinaciones, pues el agente no puede verificar su propia integridad usando los mismos mecanismos potencialmente viciados (Raghavan & Schneier, 2025).

Frente a esta nueva realidad de riesgos algorítmicos, la postura tradicional de la seguridad basada en confidencialidad, integridad y disponibilidad, se queda corta para explicar y tratar de avanzar en el aseguramiento de las nuevas iniciativas basadas en inteligencia artificial, donde el reto como que puede observar no está en si se tiene o no acceso a la información o si está disponible o no, sino en la integridad de la información, más allá de que los datos estén completos y documentados sus cambios, sino asegurar que la interpretación de los resultados de la ejecución del modelo coincida con la intención humana o la realidad externa (Raghavan & Schneier, 2025).

En este sentido, plantear una estrategia de aseguramiento para los riesgos algorítmicos demanda retomar las reflexiones planteadas por Parker (1998), donde adicional a los principios básicos de seguridad de la información: confidencialidad, integridad y disponibilidad, se incluyan autenticidad, utilidad y posesión, los cuales resultan de especial interés en esta era de algoritmos y despliegue de agentes autónomos, que pueden terminar generando ataques no tradicionales, con efectos sistémicos que aún no se pueden determinar.

Por tanto, este artículo plantea una introducción al tratamiento de los nuevos riesgos algorítmicos, aquellos que se derivan del uso, despliegue o explotación maliciosa de sis-

temas de Inteligencia Artificial (IA), incluyendo la IA generativa (IA Gen) y los agentes autónomos (IA Agéntica), que permita retar los fundamentos actuales de la seguridad de la información, para plantear algunas propuestas de seguridad y control, y así abordar el desafío de aseguramiento que implica la acelerada implementación y puesta en operación de soluciones de agentes y algoritmos basados en *machine learning*.

Confidencialidad, integridad y disponibilidad: ¿por qué ya no son suficientes en la era de la IA?

Los fundamentos de seguridad de la información nacen en una era de flujos de información, control de acceso y tensiones internacionales por cuenta de la guerra fría. En esa época el reto se traducía en el ejercicio de confidencialidad: información correcta a personas correctas, con el fin de mantener la tranquilidad de las diferentes partes interesadas e involucradas (2011). En esa época, la información como fundamento del ejercicio del poder, configuraba las relaciones a nivel internacional para lograr una posición estratégica, que le permitía a las naciones y sus aliados tomar acciones de forma anticipada con la mayor afectación y la mínima capacidad de respuesta de su contraparte.

De igual forma, mantener un registro de la información obtenida en medio de las confrontaciones y la inteligencia entre los diferentes actores en disputa, genera la necesidad de asegurar su disponibilidad de forma oportuna y confiable, con el fin de elaborar los informes requeridos para la toma de decisiones, y cruzar diferentes tipos de fuentes de información. En este contexto, las bases de datos aparecen como elementos claves, no sólo para almacenamiento ordenado y efectivo de los datos, sino como la configuración de una fuente centralizada y con control de acceso estricto para generar los análisis de información por las personas autorizadas.

Si bien la integridad de la información en esa época era un elemento importante, muy pocos avances se dieron en aquellos momentos y aún hoy, la deuda con la integridad se sigue acumulando con la llegada y evolución acelerada de la inteligencia artificial. 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 (Parker, 1998).

En términos prácticos, el objetivo de la integridad ahora es asegurar que la información y la lógica de decisión del agente de IA no sean alteradas o corrompidas, incluso si la entrada inicial fue maliciosa. Esto es, volver resistente la arquitectura donde funciona la iniciativa a ataques de inyección de *prompts* y

mantenerse ajustado al contexto de la organización, limitando su desconexión con las reglas del negocio (Díaz et al., 2025). Lo que supone incorporar un monitor de referencia, que media y controla todas las solicitudes o entradas, bien de los usuarios u otras IA, a los recursos del sistema, aplicando la política de seguridad del sistema para asegurar que solo se realicen operaciones autorizadas.

En estén sentido, Parker (1998) anticipa la necesidad de incorporar más allá de confidencialidad, integridad y disponibilidad, tres principios adicionales que se hacen necesarios para enmarcar la nueva era de los riesgos algoritmos y empezar una visión extendida de la comprensión y tratamiento de este riesgo. Los tres principios adicionales son la autenticidad, la utilidad y la posesión los cuales se describen en conjunto con los tradicionales con un ejemplo aplicado al despliegue de un agente de IA (Ver tabla No.1).

Tabla 1. Aplicación de los seis principios de seguridad de Parker

Elemento	Descripción para el Agente
Disponibilidad	El agente debe **acceder al inventario en tiempo real** a través de la API (*Application Program Interfase*) del sistema de gestión. Si la red falla o la base de datos se cae, la disponibilidad se pierde.
Utilidad	El agente necesita que los datos del inventario estén en un **formato legible y actual** (por ejemplo, un recuento numérico preciso, no un archivo de registro corrupto) para tomar decisiones de pedido. Si el agente solo puede recuperar la información de inventario de hace una semana o en un formato incomprensible (una pérdida de utilidad), la decisión de reabastecimiento será errónea.
Integridad	El agente debe asegurar que los documentos de origen, una vez recuperados de la base de conocimiento (RAG – *Retrieval-Augmented Generation*), estén **completos y que no hayan sido modificados** internamente desde su almacenamiento.
Autenticidad	El agente debe confirmar que los documentos recuperados son **genuinamente la última versión válida** firmada por la autoridad correcta (conformidad con la realidad). Si un atacante reemplaza un informe financiero real por uno falso, pero con el formato correcto, la autenticidad se pierde por **manipulación**, incluso si el archivo falso tiene una "integridad" interna asegurada (está completo).
Confidencialidad	El agente debe asegurar que la información personal contenida en el correo electrónico **no sea revelada** ni **observada** por partes no autorizadas. Esto se viola si el agente incluye accidentalmente esta información en una respuesta genérica que luego es vista por el público o por otros empleados sin autorización.
Posesión	El agente y, por extensión, la organización, deben mantener el **control físico y lógico** sobre la información del cliente. 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, ya que han pasado al control de un adversario. La pérdida de control (posesión) es un riesgo directo de las acciones deshonestas de los agentes.

Nota: Elaboración propia con ideas de Parker, 1998.

Riesgos algorítmicos. Una nueva frontera para la gestión de riesgos empresariales

Cuando se desarrolla la gestión de riesgos empresariales de forma tradicional dos retos posiblemente estén quedando fuera del radar de las organizaciones. Uno es entender que las organizaciones ahora se sitúan dentro de un ecosistema digital de negocios, por lo tanto no es sólo detallar y reconocer cómo la organización funciona de acuerdo con los retos y exigencias esperadas para cumplir con sus objetivos de negocios, sino cómo se relaciona y crea capacidades claves con sus terceros de confianza, para entregar experiencias distintas en sus clientes.

El otro es la emergencia de riesgos cognitivos, aquellos asociados con la capacidad de una persona o grupos de personas de crear contexto hiperrealistas, combinando información confiable con datos falsos para crear narrativas y situarlas en el imaginario de individuos o grupos sociales específicos aprovechándose de las vulnerabilidades sociales, cognitivas y tecnológicas (Bone & Lee, 2023), y de riesgos algorítmicos, previamente definidos.

El riesgo algorítmico es la posibilidad de un daño, pérdida financiera o afectación de la reputación de una organización que surge por: (Godhrawala, 2025)

- *Fallos internos del sistema de IA*: Sistemas que operan como "cajas negras" y cuyas decisiones son difíciles de interpretar y auditar. Estos fallos pueden generarse debido a comportamientos impredecibles (agentes autónomos que se desvían de los objetivos previstos), amplificación de sesgos presentes en los datos de entrenamiento que conducen a resultados discriminatorios, o errores que resultan en fallas operacionales y sobrecarga de recursos.

- *Explotación por actores maliciosos*: El uso de IA por parte de adversarios (cibercriminales, crimen organizado, personas inescrupulosas o actores no estatales y estatales) para automatizar, acelerar y escalar las amenazas cibernéticas. 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.

El riesgo algorítmico introduce vulnerabilidades novedosas e inéditas que no existían en los sistemas tradicionales. La IA Agéntica y la IA Gen al operar en entornos inherentemente hostiles e interactuar con fuentes no confiables, introduce vulnerabilidades estructurales que generan retos en la interpretación de sus resultados. El problema fundamental radica en que la IA debe

comprimir la realidad en formas legibles para sus modelos, creando una brecha semántica que puede ser explotada por los adversarios (Raghavan & Schneier, 2025).

Para comprender mejor este escenario de riesgos algorítmicos se aplica el modelo OODA (Observar, Orientar, Decidir y Actuar) para los agentes de inteligencia artificial, particularmente autónomos como se observa en la tabla No.2. El modelo OODA fue introducido por el Coronel John Boyd de la Fuerza Aérea de los Estados Unidos de América hace décadas. Se concibió como un marco para que los pilotos de combate comprendieran la toma de decisiones continua en tiempo real. En este sentido, un agente de IA, al igual que un piloto, ejecuta este ciclo repetidamente para lograr sus objetivos dentro de

Tabla 2. Aplicación del modelo OODA para una IA Agéntica

Fase OODA	Definición	Riesgos de la IA Agéntica	Implicación de Seguridad Clave
Observar	Recopilación de información en tiempo real.	• Inyección de prompts. • Los atacantes proporcionan las observaciones y manipulan la salida.	• La capa de observación **carece de autenticación e integridad**. • Las instrucciones maliciosas ocultas en los datos pueden afectar el resultado.
Orientar	Construcción de la "visión del mundo" del agente basada en las observaciones.	• Envenenamiento de datos de entrenamiento • Manipulación de contexto • Puertas traseras semánticas[1]	La orientación del modelo puede ser influenciada meses antes del despliegue, activando comportamientos codificados con frases de activación.
Decidir	Formulación de un plan de acción.	• Corrupción Lógica mediante ataques de ajuste fino (*fine-tuning*) • Manipulación de recompensas • Desalineación de objetivos	El proceso de decisión probabilístico del LLM se convierte en la carga útil del ataque, y los modelos pueden ser manipulados para confiar en fuentes maliciosas preferentemente.
Actuar	Ejecución del plan de acción mediante el uso de herramientas.	• Manipulación de la salida, • Confusión de herramientas • Secuestro de Acciones	Cada llamada a una herramienta confía implícitamente en las etapas previas.

Nota: Basado en: Raghavan & Schneier, 2025

[1] Es una forma de compromiso de la integridad de un modelo de lenguaje grande (LLM) o agente de IA, que explota su proceso de aprendizaje para inyectar un comportamiento malicioso latente que solo se activa bajo condiciones o frases específicas conocidas como disparadores (*triggers*) (Chen et al., 2025)

un entorno en constante cambio. Los sistemas de IA Agéntica son sistemas diseñados para percibir su entorno, tomar decisiones y ejecutar acciones autónomas para alcanzar metas definidas por el usuario (Raghavan & Schneier, 2025).

Auditoría algorítmica. Nueva práctica de seguridad y control en la era de la IA

La Auditoría algorítmica (o de IA) previa al despliegue de una iniciativa de IA es 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 (Godhrawala, 2025).

El objetivo central es prevenir la materialización de riesgos algorítmicos (como sesgos, fallas operacionales, o inyección de instrucciones maliciosas) mediante la evaluación detallada del diseño, los datos de entrenamiento y los controles de gobierno y seguridad, antes de que el sistema comience a tomar decisiones de forma autónoma.

Para aplicar una auditoría algorítmica, es necesario enfocarse en los aspectos que aseguren la confianza de los resultados, la resiliencia frente a fallas y el cumplimiento regulatorio vigente a la fecha (por ahora fragmentado y poco claro): (Mckinsey, 2025).

1. Gobernanza y Responsabilidad (*Accountability*):

- Matriz de responsabilidad: Se debe evaluar si existe un equipo de gobernanza interdisciplinario y si se han desarrollado matrices de responsabilidad claras que definan la rendición de cuentas por las acciones de cada agente.

- Alineación estratégica: Asegurar que la estrategia de IA esté alineada con los objetivos de negocio y de Recursos Humanos, y que los procesos para mantener esta alineación existan y operen.

- Controles de terceros: Verificar que los procesos de adquisición de IA de terceros son robustos y que se ha negociado el derecho de auditar los modelos o servicios contratados.

2. Transparencia y Explicabilidad:

- Modelos interpretables: Confirmar que los sistemas, especialmente aquellos que producen resultados no matemáticos, despliegan modelos de IA Explicable que permiten interpretar y auditar la lógica de la decisión.

- Registros Inmutable: Asegurar que el sistema está diseñado para generar rastros de auditoría inmutables (no alterables) que capturen todas las acciones del

agente para asegurar la rendición de cuentas y la trazabilidad de los errores.

3. Riesgo de Datos y Sesgo:

- Calidad de datos: Auditar la calidad y curaduría de los conjuntos de datos de entrenamiento.

- Mitigación de sesgos: Auditar sistemáticamente los datos de entrenamiento para verificar su confiabilidad, y así limitar que la IA perpetúe sesgos presentes en los datos, que generen resultados discriminatorios.

- Protección de datos: Verificar que se hayan implementado políticas de clasificación de datos y que se proteja la propiedad intelectual y los datos propietarios utilizados en el entrenamiento.

4. Seguridad por Diseño y Pruebas:

- Analítica de comportamiento: Asegurar que existen capacidades para monitorear el comportamiento y los procesos de toma de decisiones de los modelos a lo largo de su ciclo de vida.

- Detección de novedad: Verificar la capacidad del modelo para detectar entradas que se encuentran fuera de su dominio de competencia o experiencia.

La auditoría algorítmica exige un conjunto de habilidades híbridas que combinan el conocimiento tecnológico profundo con la visión de gobernanza, riesgo y estrategia empresarial. En este sentido, el auditor moderno deberá desarrollar una serie de competencias claves para adelantar este nuevo tipo de auditorías, detalladas en la tabla No.3.

Riesgos algorítmicos. Repensar los marcos de seguridad y control

Los riesgos algorítmicos crean un escenario novedoso para las organizaciones, pues el reto no sólo es seguridad y privacidad por diseño, sino crear un entorno de confianza digital que ofrezca condiciones de ejecución confiables, y el adecuado tratamiento de la información, para asegurar la rendición de cuentas por uso y despliegue de agentes con IA.

Lo anterior pasa por temas como:

- Incorporar interruptores de parada y sistemas de verificación multi-agente.

- Asegurar un monitoreo continuo (24x7) y afinamiento del SIEM (*Security Information and Event Management*) para detectar anomalías en tiempo real y responder en segundos.

- Instalar un control de acceso estricto y múltiple factor de autenticación para todos los agentes y servicios.

- Incluir derechos de auditoría en contratos de proveedores de IA/SaaS (*Software as a Service*).

Tabla 3. Competencias claves para la auditoría algorítmica

Competencia	Requisitos clave para Auditoría Algorítmica	Aplicación Específica en IA
Fluidez técnica y de datos	Comprender en profundidad la tecnología y sus limitaciones.	• **Alfabetización en IA**: Entender LLMs, GenAI y la arquitectura de agentes. • **Explicabilidad de la AI**: Desplegar modelos interpretables para auditar lógicas de decisión opacas ("cajas negras"). • **Datos**: Asegurar gobernanza de datos, calidad y trazabilidad.
II. Gobernanza, Riesgo y Cumplimiento (GRC)	Asegurar marcos de control que se adapten a la toma de decisiones autónoma.	• **Gobernanza evolutiva**: Establecer controles para la trazabilidad, la observabilidad y la seguridad adaptativa. • **Rendición de cuentas**: Definir matrices de responsabilidad para agentes. • **Supervisión humana**: Requerir intervención humana en decisiones de alto riesgo.
III. Ciberseguridad y riesgos emergentes	Entender y anticipar el *crimen algorítmico* y amenazas disruptivas.	• **Inteligencia de amenazas basadas en IA**: Analizar riesgos de *deepfakes*, *phishing* generado por IA, y plataformas *Crime-as-a-Service*. • **Observabilidad**: Asegurar que las herramientas detecten anomalías conductuales de agentes. • **Switch de apagado**: Integrar desde el diseño el botón de apagado del agente.
IV. Habilidades humanas y estratégicas	Liderar la adaptabilidad y construir confianza frente a la incertidumbre.	• **Adaptabilidad:** Priorizar el aprendizaje continuo para igualar el ritmo de la amenaza. • **Anticipación:** Usar la planificación de escenarios para abordar riesgos sistémicos. • **Ética:** Actuar como "Arquitecto de la Confianza Digital", asegurando el uso ético y transparente de la IA.

Nota: Elaboración propia basado en: IIA, 2025; PwC, 2025

- Implementar módulos de IA Explicable y trazabilidad inmutable (registros de logs del procesos para llegar a los resultados no modificables) para asegurar la auditabilidad y el cumplimiento regulatorio.
- Simular ataques de *deepfake* y desinformación generada por IA en ejercicios de la junta directiva.
- Hacer una extensión del seguro cibernético para los posibles daños y afectaciones de los agentes desplegados.
- Promover formación en alfabetización digital sobre los agentes

de inteligencia artificial sus ventajas y limitaciones para el equipo directivo.

- Requerir la aprobación humana para decisiones de alto riesgo que generen los agentes de inteligencia artificial.

- Entender los componentes técnicos, cómo los modelos grandes de lenguaje (LLMs) y las arquitecturas de agentes perciben, orquestan y actúan de forma autónoma.

Como se puede observar, no es solamente situarse en los principios tradicionales de la seguridad, sino establecer una vista holística que integre la seguridad y la privacidad desde el diseño, la ética, la equidad y la explicabilidad, para abordar, no sólo la defensa del modelo de la IA y sus datos, sino el impacto social y la confiabilidad a largo plazo. En pocas palabras, se debe configurar una "confianza por diseño - CpD" que asegure la rendición de cuentas de la organización en el despliegue de sistemas complejos (ahora basados en agentes con IA) y la supervisión vigilante de su desempeño ético (Welle, 2025).

El CpD no sustituye a la seguridad por diseño o la privacidad por diseño, sino que actúa como un marco de gobernanza superior que los engloba y los extiende para abordar los desafíos de la IA.

A continuación un resumen consolidado de las tres perspectivas actuales alrededor de la seguridad y control con sus definiciones y limitaciones (Ver tabla Tabla 4).

Tabla 4. Perspectivas actuales de seguridad y control

Perspectiva	Definición	Limitaciones
Seguridad por diseño (SpD)	Protección de la infraestructura y el sistema contra amenazas. Integración de controles técnicos (cifrado, autenticación, etc).	No aborda el impacto social de un sistema técnicamente confiable. Un algoritmo puede ser confiable pero éticamente sesgado.
Privacidad por diseño (PpD)	Protección del dato personal. Se enfoca en los derechos del interesado sobre su información.	Es insuficiente cuando el riesgo proviene del uso del modelo (ej. decisiones discriminatorias) y no del dato en sí, especialmente si usa datos anónimos o no personales.
Confianza por diseño (CpD)	Enfoque holístico. Integra SpD, PpD, ética, equidad y explicabilidad. Aborda el impacto social y la confiabilidad a largo plazo.	Asegurar una vista integrada del reto de la implementación de la inteligencia artificial a nivel empresarial que incluya la alineación con el negocio, el apetito de riesgo cibernético empresarial y sus impactos (oportunidades y amenazas)

Nota: Elaboración propia con ideas: Behbahani, 2025

Conclusiones

La IA ya no es una promesa tecnológica, sino un motor de cambio que impulsa la automatización de flujos de trabajo complejos y la toma de decisiones. Dado que los sistemas de IA, especialmente la IA Agéntica, son capaces de tomar decisiones y ejecutar acciones de forma autónoma, y su complejidad puede hacerlos "cajas negras", la función de auditoría interna debe evolucionar para asegurar la confiabilidad, la transparencia y la rendición de cuentas tanto para la empresa como para el ecosistema digital donde opera.

En este sentido, el riesgo algorítmico se configura como un acelerador significativo de amenazas cibernéticas que enfrentan las organizaciones hoy, que hace evidente su característica sistémica.

La rápida adopción de la IA Generativa (IA Gen) y los agentes autónomos (IA Agéntica) está redefiniendo el panorama de riesgos, llevando la ciberseguridad de una comprensión exclusivamente técnica a un imperativo estratégico que afecta la estabilidad financiera, la reputación y la continuidad del negocio.

El marco tradicional de seguridad basado en confidencialidad, integridad y disponibilidad (CID) se considera incompleto y limitado para los desafíos modernos. Parker (1998) propone un nuevo marco de análisis que extiende los elementos mencionados de seguridad y control incluyendo ahora la utilidad, la posesión y la autenticidad, los cuales, para el momento actual con la incorporación de la IA Gen y sistema autónomos, resultan del mayor interés la combinación entre integridad y autenticidad.

Los sistemas de IA Generativa y los Agentes de IA son inherentemente no deterministas, lo que significa que la misma *entrada* puede generar una variedad de posibles *salidas*, haciéndolos difíciles de gestionar y vulnerables a errores. Los principales riesgos que amenazan directamente la autenticidad y la integridad incluyen: alucinaciones, inyección de *prompts*, envenenamiento de datos, divulgación de datos sensibles y degradación del modelo, los cuales hacen evidente el reto de la integridad semántica que no le permite al agente interpretar la información de manera correcta, válida y conforme a la realidad.

En resumen, es necesario gestionar el riesgo algorítmico en una realidad cambiante como la actual, donde los sistemas de IA configuran un ecosistema dinámico sujeto a degradación intrínseca y ataques continuos que explotan la falta de separación entre instrucciones y datos. Por tanto, 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, ya que el enfoque de ser "rápido" e "inteli-

gente" sin verificación de la información (Raghavan & Schneier, 2025), sacrifica inherentemente la seguridad y la privacidad, afectando en el mediano y largo plazo el reto real de una empresa en el contexto digital con su cliente: la "confianza por diseño".

Referencias

Behbahani, A. (2025). *Why Trust Cannot Be an Afterthought*. TÜV SÜD. https://www.tuvsud.com/en-us/resource-centre/blogs/business-assurance/trust-by-design-the-value-driven-route-to-trusted-and-certified-ai-with-iso-iec-42001

Bone, J. & Lee, J. (2023). *Cognitive risk*. Boca Raton, Fl. USA. CRC Press.

Cankaya, E. C. (2011). *Bell-LaPadula confidentiality model.* En Encyclopedia of Cryptography and Security (pp. 71–74). Springer US.

Chen, K., Zhou, X., Lin, Y., Su, J., Yu, Y., Shen, L., & Lin, F. (2025). *A survey on data security in large Language Models.* En arXiv [cs.CR]. https://doi.org/10.48550/ARXIV.2508.02312

Díaz, S., Kern, C. & Olive, K. (2025). *Google's Approach for Secure AI Agents.* Google Research. https://research.google/pubs/an-introduction-to-googles-approach-for-secure-ai-agents

Godhrawala, A. (2025). *How Agentic AI can transform industries by 2028.* EY Agentic AI Series. https://www.ey.com/en_in/insights/ai/how-agentic-ai-can-transform-industries-by-2028

Harrison, M., Ruzzo, W. & Ullman, J. (1976). *Protection in operating systems*. Communications of ACM. *19*(8). https://doi.org/10.1145/360303.360333

IIA (2025). *Risk in focus. Hot topics for internal auditor 2026.* IIA. https://www.theiia.org/globalassets/site/foundation/latest-research-and-products/risk-in-focus/2026/2026-global-report-en-riskinfocus.pdf?v=0925202501

Janapa Reddi, V. (2025). *Introduction to machine learning systems. Principles and Practices of Engineering Artificially Intelligent Systems*. School of Engineering and Applied Sciences. Harvard University. https://www.mlsysbook.ai/assets/downloads/Machine-Learning-Systems.pdf

Mckinsey (2025). Agentic AI security: Risks & governance for enterprises. https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/deploying-agentic-ai-with-safety-and-security-a-playbook-for-technology-leaders

Parker, D. (1998). *Fighting computer crime. A New Framework for Protecting Information. New York*, NY. USA: John Wiley & Sons.

PwC (2025). Global digital trust insights 2026. https://www.pwc.com/us/en/services/consulting/cybersecurity-risk-regulatory/library/global-digital-trust-insights.html

Raghavan, B., & Schneier, B. (2025). *Agentic AI's OODA Loop Problem*. IEEE security & privacy, 2–4. https://doi.org/10.1109/msec.2025.3604105

Welle, J. (2025). *From compliance to competitive advantage: Building trust by design*. Capgemini.

https://www.capgemini.com/insights/expert-perspectives/from-compliance-to-competitive-advantage-building-trust-by-design

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–.

Inteligencias Paralelas

Puentes entre las redes neuronales biológicas y las redes neuronales computacionales

DOI: 10.29236/sistemas.n177a7

Resumen

La relación entre las redes neuronales biológicas y las redes neuronales artificiales constituye uno de los diálogos científicos más fértiles de la actualidad. Aunque surgieron en contextos distintos, ambos sistemas buscan comprender y modelar la inteligencia desde perspectivas complementarias: la biología mediante la exploración de sinapsis, plasticidad y circuitos funcionales, y la computación mediante arquitecturas matemáticas capaces de aprender a partir de datos. Este artículo examina los principios fundamentales de cada enfoque y analiza sus puentes conceptuales, evitando reducir uno al otro. Se abordan temas como la organización del conectoma, la dinámica sináptica, la evolución histórica de las redes artificiales y su capacidad para procesar información mediante optimización estadística. Asimismo, se exploran escenarios de convergencia como las interfaces cerebro–computador, la computación neuromórfica y las analogías emergentes entre el olvido biológico y los métodos de *machine unlearning*. A través de este análisis, se muestra que la interacción entre neurociencia e inteligencia artificial no persigue imitación estructural, sino inspiración funcional y expansión mutua. La comparación crítica entre ambos sistemas permite comprender mejor sus alcances, límites y posibilidades, y abre nuevas vías para el desarrollo de tecnologías que complementen, y no sustituyan, las capacidades humanas.

Palabras claves

Neurociencia, Inteligencia Artificial, Redes Neuronales, Plasticidad Sináptica, *Machine Unlearning*

Introducción

"Lo que hoy llamamos inteligencia artificial no es más que el esfuerzo por entender la inteligencia humana."
— Marvin Minsky, uno de los fundadores de la IA.

La búsqueda por comprender la inteligencia ha conducido, desde mediados del siglo XX, a un diálogo cada vez más estrecho entre neurociencia y ciencias de la computación. Dos trayectorias que nacieron separadas —una enfocada en desentrañar la arquitectura viva del cerebro y otra en construir modelos matemáticos capaces de aprender— han confluido gradualmente en un terreno común. La neurociencia ha revelado que la mente humana es un sistema profundamente interconectado, organizado en redes y subredes cuya dinámica puede describirse mediante principios de conectividad estructural y funcional (Bassett & Sporns, 2017). Al mismo tiempo, la inteligencia artificial moderna se ha expandido desde modelos simples como el perceptrón (McCulloch & Pitts, 1943; Rosenblatt, 1958) hacia arquitecturas profundas capaces de interpretar imágenes, producir lenguaje y resolver tareas complejas en dominios ampliamente variables (LeCun et al., 2015).

En las últimas décadas, logros como la victoria de *AlphaGo* sobre el campeón mundial de Go, el desempeño sobresaliente de agentes computacionales en videojuegos de estrategia en tiempo real y los avances de los vehículos autónomos han mostrado que las máquinas pueden aprender patrones sutiles sin depender de reglas explícitas. Estos sistemas están construidos sobre redes neuronales profundas entrenadas con grandes volúmenes de datos, utilizando principios matemáticos que, aunque inspirados remotamente en el cerebro, operan bajo lógicas muy distintas (Goodfellow et al., 2016). En lugar de reproducir procesos biológicos, estas redes implementan optimizaciones numéricas que permiten ajustar millones de parámetros para lograr comportamientos adaptativos.

Mientras la informática avanza en su propio marco conceptual, la neurociencia sigue describiendo el cerebro con un nivel de detalle cada vez más fino. El conocimiento actual sobre la sinapsis, los mecanismos de potenciación a largo plazo y depresión sináptica, y las bases moleculares del aprendizaje revelan un sistema plástico, químico y altamente especializado (Kandel et al., 2013). Las investigaciones sobre el conectoma y la organización en redes funcionales demuestran que el procesamiento cognitivo emerge de la interacción coordinada de regiones distantes, en esca-

las temporales que van de los milisegundos a varios segundos (Edlow & Menon, 2024). A diferencia de las redes artificiales, cuya arquitectura es diseñada y definida de antemano, las redes biológicas se reorganizan permanentemente en función de la experiencia y el entorno.

A pesar de sus diferencias fundamentales, ambos campos convergen en conceptos que resuenan de manera profunda. Tanto las redes biológicas como las artificiales operan mediante conexiones entre unidades elementales, y en ambos casos el aprendizaje surge de la modificación de esas conexiones. Pero la semejanza es conceptual, no estructural. Las neuronas biológicas son entidades complejas que integran señales electroquímicas, participan en mecanismos oscilatorios y dependen del soporte de células gliales; las unidades artificiales, en contraste, son funciones matemáticas sencillas que reciben entradas numéricas, aplican transformaciones y generan salidas. Allí donde el cerebro utiliza neurotransmisores, ritmos oscilatorios, neuromodulación y un entorno químico dinámico, las redes artificiales utilizan ecuaciones lineales y funciones de activación como ReLU o sigmoide.

La comparación se vuelve aún más rica cuando se examina el papel del olvido. En neurociencia, el olvido cumple funciones adaptativas esenciales: evita la saturación del sistema, optimiza la flexibilidad cognitiva y permite reorganizar el conocimiento en función de nuevas experiencias. Procesos como la poda sináptica en el desarrollo o la depresión sináptica en circuitos maduros ilustran cómo el cerebro elimina conexiones para mejorar su eficiencia (Kandel et al., 2013). En un plano completamente distinto, pero conceptualmente cercano, la inteligencia artificial contemporánea ha comenzado a explorar el machine unlearning, un conjunto de técnicas que buscan permitir a los modelos eliminar información previamente aprendida sin reconstruirlos desde cero. Esta aproximación surge de la necesidad de cumplir requisitos de privacidad y de corregir sesgos o dependencias indeseadas en el entrenamiento, y representa un desafío técnico aún abierto (Xu et al., 2023). Aunque la analogía es limitada, resulta llamativo que la computación esté empezando a enfrentar —desde otra lógica— problemas que la biología resolvió hace millones de años.

La intersección entre neurociencia e inteligencia artificial no pretende fusionar ambos dominios, sino permitir que se iluminen mutuamente. La biología aporta modelos de organización complejos, eficientes y sorprendentemente robustos; la computación ofrece abstracciones matemáticas que permiten explorar hipótesis, visualizar patrones y diseñar sistemas adaptativos. El diálogo entre ambos campos abre la puerta a nuevas preguntas sobre el

aprendizaje, la memoria, el olvido y la representación de la información, y fortalece la comprensión de dos sistemas que, aunque profundamente distintos, comparten la aspiración de explicar cómo emerge la inteligencia.

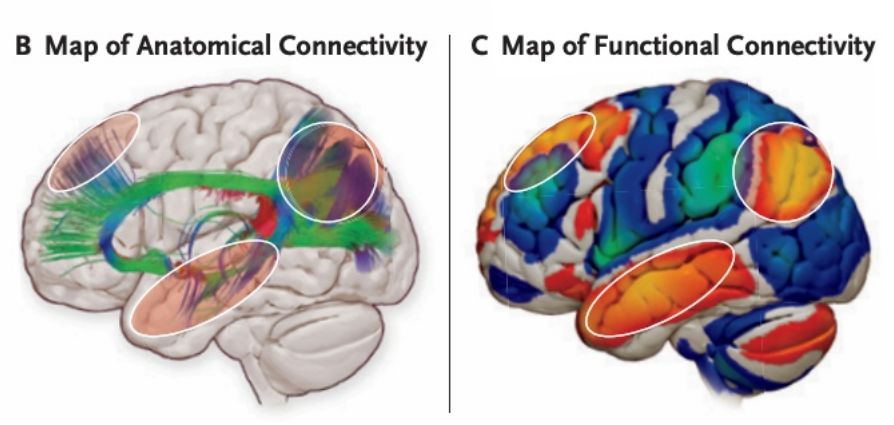
Fundamentos Biológicos de las Redes Neuronales Naturales (RNB)

Comprender la organización del cerebro humano es adentrarse en uno de los sistemas más complejos que existen en la naturaleza. Su arquitectura, formada por aproximadamente ochenta y seis mil millones de neuronas y billones de sinapsis, no sigue un diseño estático ni rígido: evoluciona, se reorganiza, se fortalece y se debilita en función de la experiencia. Mientras que las redes artificiales se construyen a partir de ecuaciones y parámetros predefinidos, las redes biológicas emergen de interacciones bioquímicas y electrofisiológicas cuya riqueza desafía cualquier abstracción matemática.

La unidad fundamental de comunicación neuronal es la sinapsis. Lejos de ser un simple punto de contacto entre neuronas, la sinapsis constituye un espacio altamente especializado en el que convergen señales químicas y eléctricas que modelan el flujo de información. Existen sinapsis eléctricas —más rápidas, basadas en uniones gap— y sinapsis químicas —más lentas pero modulables—, además del modelo sináptico tripartito donde intervienen los astrocitos, añadiendo una capa adicional de regulación (Kandel et al., 2013). La presentación Redes Neuronales Biológicas ilustra cómo la sinapsis es mucho más que una conexión: es un microambiente que sincroniza, amplifica o inhibe información de forma dinámica.

A partir de estas unidades microscópicas emergen estructuras mesoscópicas como las columnas corticales, bloques funcionales que organizan el procesamiento sensorial y cognitivo. Estas columnas no son entidades aisladas, sino nodos altamente interconectados en redes distribuidas. En regiones como la corteza sensorial primaria, por ejemplo, convergen señales del mundo externo que luego se distribuyen jerárquicamente hacia áreas superiores encargadas de integrar y reinterpretar esa información (Mountcastle, 1997). Es en esta integración local y global donde se esconde gran parte de la sofisticación cognitiva humana.

En un nivel aún más amplio, el conectoma cerebral constituye la arquitectura global de conexiones estructurales y funcionales que vinculan regiones distantes del cerebro. Estudios de neuroimagen han permitido observar cómo estas redes presentan patrones robustos de conectividad, organizados en módulos funcionales que interactúan mediante hubs que actúan como centros logísticos de información



Fox MD. N Engl J Med. 2018 Dec 6;379(23):2237-2245

(Sporns, 2011). Dichas redes presentan propiedades de eficiencia, resiliencia y organización jerárquica que han inspirado modelos computacionales de procesamiento distribuido (Bassett & Sporns, 2017).

El sistema visual es uno de los ejemplos más elocuentes de organización jerárquica en el cerebro. Señales provenientes de la retina viajan al tálamo y luego a la corteza visual primaria, donde neuronas especializadas responden a líneas, orientaciones o movimientos específicos. A medida que la información asciende hacia áreas superiores, se combinan características simples hasta construir representaciones complejas como rostros, objetos o escenas (Kandel et al., 2013). Esta lógica jerárquica recuerda, en cierta medida, la manera en que las redes convolucionales procesan imágenes, aunque el paralelismo es conceptual y no estructural.

Otro ejemplo es la red de memoria episódica que vincula el hipocampo con la corteza prefrontal. Esta red permite registrar, consolidar y recuperar experiencias personales y, al mismo tiempo, suprime información irrelevante para evitar la saturación. Su funcionamiento muestra que la memoria biológica no es un archivo estático, sino una reconstrucción dinámica y en constante actualización, afectada por emociones, contexto y expectativas.

A gran escala, las redes funcionales identificadas mediante resonancia magnética funcional —como la red por defecto, la red frontoparietal o la red sensoriomotora— revelan cómo el cerebro coordina regiones distantes para ejecutar tareas complejas como la introspección, la planificación o la atención sostenida (Fox, 2018). Estas redes funcionan como sistemas cooperativos donde la activación de una región tiene repercusiones sistémicas en muchas otras.

A pesar de su sofisticación, el cerebro humano no es un procesador perfecto. Su actividad está atravesada por ruido neuronal, variabilidad intrínseca y limitaciones energéticas. Este ruido, lejos de ser un problema, favorece la adaptabilidad y evita el determinismo rígido, permitiendo que el sistema explore múltiples configuraciones funcionales (Faisal et al., 2008). De hecho, mientras que los grandes modelos de IA consumen cantidades masivas de energía, el cerebro opera con apenas unos 20 vatios, una eficiencia que aún no tiene paralelo artificial.

Su naturaleza electroquímica, aunque genera un inmenso repertorio de posibilidades, introduce también límites: las señales tardan milisegundos en propagarse, la reorganización sináptica requiere tiempo biológico y la plasticidad está influenciada por factores hormonales, emocionales y genéticos. Lejos de ser una desventaja, esta lentitud relativa permite que el cerebro priorice estabilidad, aprendizaje contextual y resiliencia.

Finalmente, el cerebro es producto de millones de años de evolución, no de diseño racional. Su arquitectura refleja una historia de adaptaciones, compromisos y reutilización de estructuras que, aunque eficiente, no sigue los principios de optimización computacional que orientan la ingeniería moderna (Edelman, 1987). Esta diferencia epistemológica es clave para entender por qué compararlo con una red artificial puede ser útil metafóricamente, pero insuficiente en términos estructurales.

Fundamentos Computacionales de las Redes Neuronales Artificiales

La idea de construir modelos computacionales inspirados en el cerebro no es nueva; surgió en el corazón de la cibernética, cuando matemáticos y neurofisiólogos empezaron a preguntarse si los principios de la actividad neuronal podían formalizarse en ecuaciones. Esa intuición inicial dio origen a un recorrido fascinante. Algunas metáforas sobrevivieron, otras se transformaron o se abandonaron, pero el objetivo persistió: diseñar sistemas capaces de aprender. Las redes neuronales artificiales, tal como se conocen hoy, son el resultado acumulado de décadas de experimentación científica, avances tecnológicos y revoluciones conceptuales.

El primer intento formal de capturar el comportamiento de una neurona en términos matemáticos fue el trabajo de McCulloch y Pitts (1943), quienes describieron una “neurona lógica”: una unidad que integraba entradas binarias y generaba una salida basada en reglas formales.

Aunque extraordinariamente simple, este modelo inauguró la posibilidad de representar procesos cognitivos mediante circuitos computacionales, Ver Figura 1.

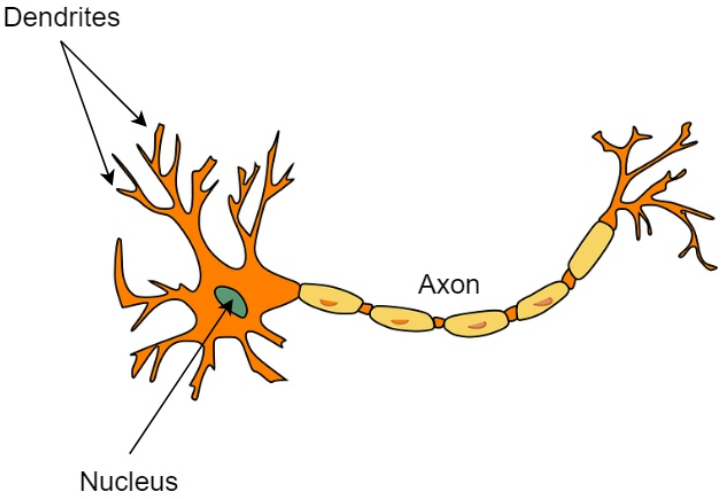


Figura 1. Anatomía de una neurona. Adaptado de Towards Data Science. https://towardsdatascience.com/the-concept-of-artificial-neurons-perceptrons-in-neural-networks-fab22249cbfc

Años después, Rosenblatt (1958) propuso el perceptrón, una estructura más flexible compuesta por pesos ajustables que permitían "aprender" relaciones lineales entre datos. Su idea, revolucionaria en su época, abrió la puerta a los sistemas adaptativos. Sin embargo, las limitaciones del perceptrón se hicieron evidentes pronto: no podía capturar relaciones no lineales, y su potencial parecía, en ese momento, estrecho. Durante un tiempo, esto produjo un estancamiento conceptual conocido como el "invierno de las redes neuronales".

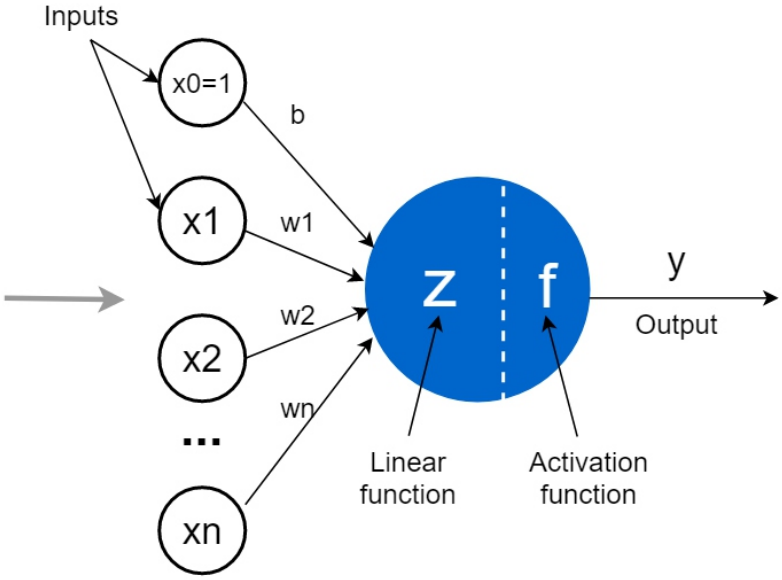
El renacimiento llegó cuando se introdujo el algoritmo de retropropagación del error en la década de 1980 (Rumelhart et al., 1986), que permitió entrenar redes de múltiples capas. Este avance transformó el campo por completo. De repente, los modelos podían resolver problemas complejos, ajustar millones de parámetros y descubrir representaciones internas sin supervisión explícita. La presentación Inteligencias Paralelas retoma ese hilo histórico y lo conecta con ejemplos contemporáneos de aprendizaje automático, mostrando cómo ese salto conceptual permitió el desarrollo de redes profundas utilizadas en visión por computadora, reconocimiento de lenguaje y análisis de grandes volúmenes de datos. Hoy, arquitecturas como las redes convolucionales, las redes recurrentes, los transformadores y los modelos generativos constituyen la columna vertebral de muchos sistemas de inteligencia artificial.

Aunque la terminología toma prestados conceptos de la neurobiología, los elementos que conforman una red artificial pertenecen al dominio estricto de las matemáticas y la estadística. Una unidad artificial recibe entradas numéricas, las multiplica por pesos, suma un sesgo y aplica una función de activación —una operación aparentemente trivial, pero fundamental para introducir no linealidad—. Esta combinación, repetida millones de veces, es la responsable de la riqueza expresiva de los modelos contemporáneos (Goodfellow et al., 2016). Ver Figura 2.

Los pesos representan la importancia relativa de cada señal de entrada. El sesgo permite desplazar la función de activación, modificando su sensibilidad. Finalmente, la activación determina la transformación aplicada: puede ser sigmoidal, hiperbólica, lineal o basada en unidades rectificadas como ReLU, ampliamente utilizadas por su eficiencia computacional.

Cuando varias de estas unidades se organizan en capas, se forma una arquitectura que procesa información de manera jerárquica. Cada capa detecta patrones de dife-

Figura 2. Diagrama de un perceptrón artificial. Adaptado R. Pramoditha (2021)



Nota: perceptrón artificial. Adaptado de The concept of artificial neurons (perceptrons) in neural networks por R. Pramoditha (2021), Towards Data Science. https://towardsdatascience.com/the-concept-of-artificial-neurons-perceptrons-in-neural-networks-fab22249cbfc

rente complejidad: las iniciales captan características simples; las intermedias, patrones abstractos; las finales, decisiones de mayor orden.

Así funciona el procesamiento en redes convolucionales aplicadas a visión por computador, donde filtros sucesivos extraen gradualmente bordes, contornos, formas y finalmente objetos completos.

El aprendizaje en redes artificiales no se basa en mecanismos electroquímicos ni en sinapsis moduladas por neurotransmisores. En cambio, se fundamenta en la optimización matemática. El objetivo consiste en minimizar una función de pérdida que mide qué tan lejos está el modelo de la predicción correcta. Para lograrlo, se emplea el descenso del gradiente y sus variantes, que ajustan los pesos de forma iterativa a partir del error.

Gracias al algoritmo de retropropagación, las redes aprenden calculando cómo cada parámetro contribuyó al error total y actualizando su valor para mejorar el desempeño en el siguiente ciclo. Aunque este procedimiento se repite millones de veces, ocurre a velocidades extraordinarias gracias a las unidades de procesamiento paralelo como las GPU y, más recientemente, las TPU.

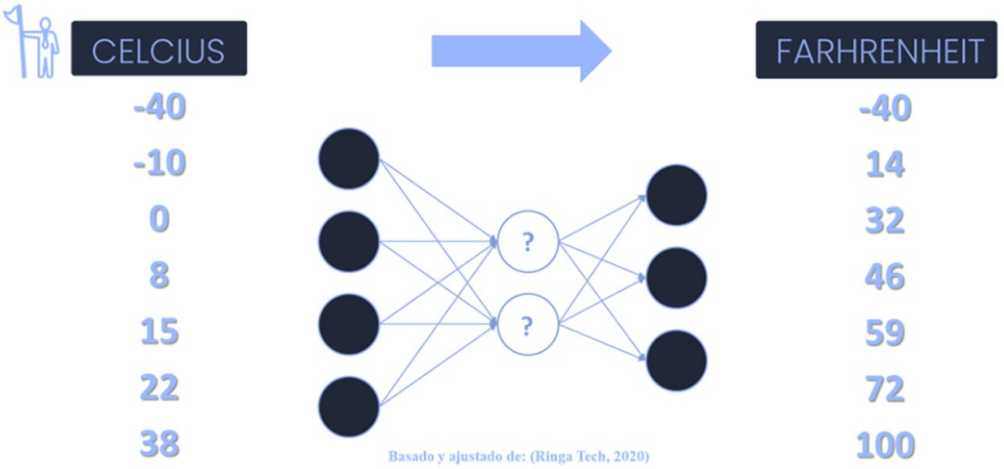
A diferencia del cerebro, que aprende incluso con pocos ejemplos y en entornos ambiguos, las redes artificiales suelen requerir grandes volúmenes de datos para obtener un rendimiento aceptable. Esta diferencia no se debe a incapacidad computacional, sino a que los modelos matemáticos carecen de los mecanismos biológicos de generalización, contexto y experiencia acumulada. El cerebro, como señalan Kandel et al. (2013), construye conocimiento sobre múltiples escalas temporales; las redes artificiales, en cambio, aprenden estadísticamente sobre representaciones distribuidas.

Un ejemplo sencillo pero poderoso: entrenar una neurona para convertir grados Celsius a Fahrenheit. Aunque la relación es lineal y no requiere de redes profundas, este caso demuestra la esencia del aprendizaje automático: el modelo no sabe la fórmula inicial, pero la descubre a través de ejemplos repetidos.

La conversión entre grados Celsius y Fahrenheit permite observar con claridad cómo difieren dos formas de resolver un mismo problema: la programación tradicional y el aprendizaje automático. En un enfoque clásico, el programador conoce la fórmula exacta ($F = 1.8C + 32$) y la codifica paso a paso.

El computador no descubre nada; simplemente ejecuta instrucciones precisas que garantizan un resultado determinista. Es la esencia del paradigma algorítmico: reglas explícitas, operaciones definidas y ausencia total de incertidumbre, Ver Figura 3.

Figura 3. Adaptación de (Ringa Tech, 2020)



Nota: Adaptación propia del autor para presentación.

En contraste, un modelo de inteligencia artificial no recibe la fórmula. Aprende a partir de ejemplos. La red neuronal observa pares de datos —por ejemplo, 10°C corresponde a 50°F— y ajusta sus pesos internos para aproximarse cada vez más a la respuesta correcta.

Este proceso de optimización mediante retropropagación, descrito originalmente por Rumelhart et al. (1986), permite que el modelo infiera la relación lineal sin haberla visto explícitamente. La presentación Inteligencias Paralelas muestra justamente cómo la red converge hacia la función correcta sin instrucciones directas, solo gracias a patrones en los datos.

El contraste es profundo: en la computación tradicional el conocimiento está en el código; en los modelos de IA, está en los datos. La máquina deja de ejecutar instrucciones y comienza a identificar regularidades, un enfoque que recuerda —de forma lejana, pero útil— la manera en que los sistemas biológicos aprenden a partir de experiencia (Kandel et al., 2013).

Los modelos modernos han evolucionado hacia estructuras colosales con capacidades sorprendentes. Las redes convolucionales son especializadas en análisis de imágenes; las recurrentes se orientan al procesamiento de secuencias; los transformadores han revolucionado el campo del procesamiento del lenguaje natural gracias al mecanismo de atención, que permite identificar qué partes de una entrada son más relevantes para una tarea específica (Vaswani et al., 2017).

Esta explosión de arquitecturas ha estado acompañada por avances en hardware y disponibilidad de datos. Pero también ha reabierto el debate sobre la naturaleza de la inteligencia: ¿son estos modelos realmente "inteligentes" o simplemente potentes optimizadores estadísticos? Investigadores como LeCun et al. (2015) muestran que la capacidad de estos modelos para captar patrones sutiles supera con creces lo que se consideraba posible hace apenas una década. No obstante, su tipo de "comprensión" dista profundamente de la cognición biológica.

Por muy avanzadas que sean, estas redes siguen dependiendo de estructuras matemáticas simples que se repiten a gran escala. Su aprendizaje es eficiente, pero estrecho; su razonamiento es profundo, pero sin comprensión semántica humana. Aun así, su valor práctico es incuestionable.

Puentes y Contrastes entre Redes Biológicas y Redes Artificiales

El diálogo entre neurociencia e inteligencia artificial se sostiene sobre un delicado equilibrio entre inspiración y distanciamiento. Las redes neuronales artificiales surgieron como una metáfora de las biológicas, pero con el tiempo evolucionaron hasta convertirse en sistemas radicalmente distintos. La neurociencia, por su parte, encontró en la IA un marco conceptual útil para imaginar cómo podrían organizarse ciertos procesos cognitivos; sin embargo, la evidencia ha demostrado que las semejanzas entre ambos sistemas son más metafóricas que estructurales. Examinar estos puentes y divergencias no solo aclara malentendidos comunes, sino que también ilumina lo que cada disciplina puede aprender de la otra.

La inspiración inicial de las redes neuronales artificiales fue explícitamente biológica. McCulloch y Pitts (1943) imaginaron una neurona como una unidad lógica capaz de integrar señales y producir una salida binaria. Décadas más tarde, el perceptrón de Rosenblatt (1958) reforzó esa intuición con mecanismos de ajuste de pesos que recordaban, superficialmente, a la plasticidad sináptica. Pero el parecido terminaba ahí.

Mientras que una neurona biológica integra procesos electroquímicos que involucran neurotransmisores, modulación glial y cascadas metabólicas (Kandel et al., 2013), una "neurona artificial" es simplemente una función matemática que combina entradas y aplica una transformación no lineal. No posee morfología, variabilidad química ni dinámica temporal compleja. Es una abstracción. Una simplificación extrema que, sin embargo, ha sido extraordinariamente útil para desarrollar sistemas de aprendizaje.

Este tipo de inspiración distante no pretende replicar el cerebro, sino capturar algunos principios funcionales: procesamiento distribuido, ajuste adaptativo de conexiones, jerarquías de información. Como señalan LeCun et al. (2015), las redes artificiales "se parecen al cerebro tanto como un avión se parece a un ave": comparten conceptos, no estructuras.

A pesar de la distancia estructural, existen paralelos conceptuales que facilitan la comprensión de ambos sistemas. La plasticidad sináptica —ya sea la potenciación a largo plazo o la depresión sináptica— puede compararse, en términos muy generales, con el ajuste de pesos en una red artificial. Ambos procesos modifican la fuerza de conexiones para mejorar el desempeño en una tarea específica.

Del mismo modo, la modularidad del cerebro, observada en su conectoma funcional, encuentra una analogía en las capas jerárquicas de las redes profundas (Bassett & Sporns, 2017). En ambos casos, distintos "módulos" se especializan en ciertos tipos de procesamiento y luego combinan su actividad para producir una salida coherente.

Estas analogías, sin embargo, deben manejarse con cautela. Una sinapsis no es un peso; una neurona no es una función de activación; una red biológica no es una arquitectura matemática. Las comparaciones funcionan como puentes pedagógicos, útiles para explicar conceptos complejos, pero insuficientes para describir las realidades profundas de cada sistema.

La divergencia entre redes biológicas y artificiales es abismal cuando se observan con detalle. El cerebro opera con una eficiencia energética extraordinaria —solo 20 vatios para sostener miles de millones de operaciones simultáneas—, mientras que los grandes modelos de IA pueden requerir megavatios de energía para entrenarse (Faisal et al., 2008). El contraste no es solo cuantitativo, sino cualitativo.

Las redes biológicas funcionan en un entorno ruidoso, variable y estocástico. Este ruido, lejos de ser un problema, facilita la adaptabilidad y la capacidad de generalización. En cambio, las redes artificiales dependen de estabilidad, precisión numérica y entornos altamente controlados. El cerebro reorganiza su conectividad en escalas temporales que van de milisegundos a meses; los modelos artificiales ajustan sus parámetros en microsegundos, pero carecen de la rica dimensión temporal que caracteriza la cognición humana.

Además, las redes artificiales aprenden patrones estadísticos sin entender su significado. Los modelos pueden identificar correlaciones extremadamente complejas, pero no poseen conciencia, intención ni comprensión contextual. La cognición humana, como señalan Kan-

del et al. (2013), está profundamente influenciada por factores emocionales, motivacionales y sociales que no tienen equivalente en la IA contemporánea.

El aprendizaje humano es un proceso multimodal que integra percepción, emoción, memoria, motivación y experiencia previa. No consiste simplemente en reconocer patrones, sino en atribuirles significado. Aprender implica interpretar el mundo a través de lentes biográficas y culturales. La IA, en cambio, aprende ajustando parámetros para minimizar errores en una tarea específica. Sus representaciones internas —por muy sofisticadas que sean— no tienen contenido semántico propio.

Este contraste se evidencia en el hecho de que los seres humanos pueden aprender con pocos ejemplos gracias a mecanismos de abstracción profunda, mientras que los modelos artificiales dependen, en general, de volúmenes enormes de datos etiquetados (Lake et al., 2017). El cerebro puede generalizar desde un caso único; las redes artificiales suelen necesitar miles.

Sin embargo, el aporte conceptual es recíproco. La IA ha permitido generar teorías sobre cómo podrían representarse ciertos cálculos cognitivos, mientras que la neurociencia ha inspirado nuevas arquitecturas. Son campos que avanzan paralelos, no en competencia, sino en diálogo.

Perspectivas de Convergencia: Neuroingeniería, IA y Cerebro

La intersección entre neurociencia e inteligencia artificial no solo revela diferencias profundas entre ambos sistemas, sino también puntos de aproximación que están abriendo horizontes tecnológicos completamente nuevos. En décadas recientes, los avances en interfaces cerebro-computador, computación neuromórfica y modelos híbridos han permitido vislumbrar escenarios donde las fronteras tradicionales entre lo biológico y lo artificial se vuelven más permeables. Esta convergencia no busca imitar al cerebro ni reemplazarlo, sino estudiar cómo sus principios pueden inspirar tecnologías más eficientes, flexibles y adaptativas. A la vez, los modelos computacionales ofrecen herramientas para reinterpretar funciones del sistema nervioso desde nuevas perspectivas.

Las interfaces cerebro–computador (BCI por su siglas en inglés) son uno de los campos donde la colaboración entre neurociencia e ingeniería ha sido más fructífera. Estos sistemas permiten traducir la actividad neuronal en comandos para dispositivos externos, posibilitando, por ejemplo, que personas con parálisis puedan mover prótesis robóticas o comunicarse mediante patrones de neuroactividad (Serruya et al., 2002). El principio es simple en concepto, pero complejo en ejecución: registrar señales cerebrales mediante electrodos invasi-

vos o no invasivos, procesarlas con algoritmos de decodificación y convertirlas en acciones digitales.

El conocimiento profundo sobre la organización cortical y las rutas motoras ha permitido desarrollar modelos de decodificación capaces de inferir intenciones de movimiento a partir de patrones neuronales. Investigaciones recientes muestran, incluso, la posibilidad de restaurar funciones sensoriomotoras mediante ciclos cerrados de retroalimentación, donde la señal decodificada genera un estímulo háptico que regresa al usuario, cerrando un circuito entre el cerebro y la máquina (Shenoy & Carmena, 2014).

Los avances en BCI plantean preguntas fascinantes: ¿puede una máquina extender las capacidades humanas? ¿Es posible integrar señales artificiales en el cerebro sin alterar su integridad funcional? Aunque estas tecnologías aún están en desarrollo, su potencial terapéutico y rehabilitador es extraordinario.

Por su parte, la computación neuromórfica surge del deseo de construir sistemas que no solo se inspiren conceptualmente en el cerebro, sino que también adopten algunos de sus principios estructurales. Chips como *TrueNorth* de IBM o *Loihi* de Intel implementan neuronas de picos (*spiking neurons*) que transmiten información en forma de eventos discretos, imitando la dinámica temporal de las neuronas biológicas (Davies et al., 2018). Estos dispositivos consumen muy poca energía y permiten ejecutar modelos con eficiencia notable, especialmente en tareas de reconocimiento sensorial y control robótico.

A diferencia de las redes tradicionales, las *spiking neural networks* (SNN) integran el tiempo como una variable fundamental, aproximándose a la forma en que las neuronas reales procesan señales. Aunque todavía están en desarrollo, representan un puente prometedor entre las limitaciones del hardware tradicional y las propiedades emergentes del cerebro humano.

El estudio de la arquitectura neural, especialmente del conectoma, ha inspirado también modelos de procesamiento distribuido que imitan la organización modular del cerebro (Bassett & Sporns, 2017). Estos modelos no buscan replicar estructuras biológicas, sino aprovechar principios como la eficiencia energética, la conectividad jerárquica y la descentralización del procesamiento.

La convergencia entre biología e inteligencia artificial abre caminos potentes, pero también inquietudes profundas. Las BCI plantean desafíos sobre privacidad neuronal: si es posible decodificar patrones de intención, ¿qué salvaguardas se requieren para evitar su uso indebido? Investigadores como Yuste et al. (2017) han propuesto marcos

éticos para proteger lo que llaman "neuroderechos", incluyendo el derecho a la identidad mental, la privacidad de la actividad cerebral y la libertad cognitiva.

En paralelo, la IA plantea dilemas sobre autonomía, responsabilidad y agencia. Los modelos actuales no comprenden el sentido de sus decisiones, pero sus resultados pueden influir en sistemas críticos como diagnóstico médico, justicia o educación. La interacción entre ambos campos exige una reflexión ética profunda sobre cómo diseñar tecnologías que amplíen la capacidad humana sin erosionar la dignidad y la autonomía.

La filosofía de la mente también encuentra un espacio en esta convergencia. Preguntas como "¿puede una máquina tener experiencias?" o "¿es la inteligencia reducible a funciones computacionales?" retoman vigor ante avances tecnológicos que, aunque no replican la cognición humana, sí la emulan en ciertos comportamientos superficiales. La diferencia ontológica entre un modelo estadístico y un cerebro consciente sigue siendo uno de los desafíos intelectuales más importantes de nuestra época.

Entre las tendencias emergentes en inteligencia artificial, pocas resultan tan conceptualmente sugerentes como el *machine unlearning*. Este campo busca permitir que los modelos eliminen información previamente aprendida sin necesidad de reconstruirse desde cero. En otras palabras, intenta que un sistema artificial "olvide" ciertos datos o patrones, ya sea para corregir sesgos, respetar solicitudes de privacidad (como el derecho al olvido) o depurar representaciones internas no deseadas (Cao & Yang, 2015; Xu et al., 2023).

Aunque el mecanismo es estrictamente computacional, guarda ecos curiosos con el funcionamiento del cerebro. El olvido biológico no es un fallo; es un mecanismo adaptativo esencial. La poda sináptica durante el desarrollo y los procesos de depresión a largo plazo (LTD por sus siglas en inglés) permiten eliminar conexiones innecesarias para mejorar la eficiencia y evitar la saturación cognitiva (Kandel et al., 2013). En el aprendizaje humano, olvidar es tan importante como recordar.

Los métodos computacionales actuales emplean técnicas como la fragmentación de datos, el reentrenamiento diferencial y la actualización localizada de parámetros para suprimir la influencia de ejemplos específicos en el modelo. En esencia, se busca revertir el rastro estadístico que ciertos datos dejaron durante el entrenamiento, imitando —de forma muy lejana, pero conceptualmente inspiradora— la capacidad del cerebro para debilitar o eliminar conexiones sinápticas.

El paralelismo no implica equivalencia. El olvido biológico es orgá-

nico, contextual, influido por emociones y experiencias. El desaprendizaje artificial es algorítmico, limitado y cuantificable. Pero ambos comparten una idea central: la eficiencia cognitiva requiere no solo adquirir información, sino también saber qué dejar atrás.

Más Allá de los Cables y las Sinapsis: Reflexiones Finales sobre Dos Inteligencias que Convergen

La comparación entre redes neuronales biológicas y redes neuronales artificiales no busca resolver un debate de superioridad, sino abrir una ventana hacia la complejidad de dos sistemas que, desde dominios radicalmente distintos, intentan responder a la misma pregunta: ¿cómo surge la inteligencia? A lo largo de este recorrido queda claro que ambos modelos —uno moldeado por millones de años de evolución y el otro construido mediante abstracciones matemáticas— iluminan diferentes aspectos del procesamiento de información. Sus coincidencias son conceptuales; sus diferencias, profundas. Sin embargo, es precisamente en ese espacio intermedio donde emergen las oportunidades más ricas.

El cerebro humano, con su plasticidad, su economía energética y su capacidad para generar significado, continúa siendo una fuente inagotable de inspiración para la ingeniería. La evidencia en neurociencia muestra que el aprendizaje no es solo un proceso de fortalecimiento sináptico, sino un equilibrio entre creación y eliminación, un sistema dinámico que redefine continuamente sus rutas internas (Kandel et al., 2013). Esa flexibilidad, difícil de capturar en términos computacionales, es esencial para explicar cómo los seres humanos pueden aprender con pocos ejemplos, adaptarse a entornos inciertos y reorganizar funciones tras una lesión.

Por otro lado, los modelos de inteligencia artificial han alcanzado niveles de desempeño que superan con creces la capacidad humana en tareas específicas. Sistemas capaces de analizar imágenes con precisión milimétrica o sintetizar lenguaje con fluidez sorprendente revelan que la eficiencia matemática puede, en ciertos dominios, compensar la carencia de comprensión semántica profunda (LeCun et al., 2015). Sin embargo, los modelos artificiales continúan siendo, en esencia, estructuras de optimización: aprenden correlaciones, no significados. Sus representaciones internas carecen de la dimensión subjetiva que caracteriza la cognición humana.

El diálogo entre ambas disciplinas no implica que la inteligencia artificial replique al cerebro, ni que el cerebro funcione como un algoritmo. Más bien, su interacción genera un intercambio fecundo: la neurociencia aporta principios como la modularidad, la plasticidad y la eficien-

cia energética; la IA ofrece modelos formales para explorar hipótesis y herramientas para interpretar patrones que antes eran invisibles. En campos como la computación neuromórfica, las interfaces cerebro–computador y el machine unlearning, esta convergencia es especialmente evidente, señalando un futuro donde lo biológico y lo artificial pueden complementarse de maneras cada vez más sofisticadas (Davies et al., 2018; Xu et al., 2023).

El futuro de esta intersección dependerá de dos factores clave. Por un lado, de la capacidad para desarrollar tecnologías que respeten la complejidad del cerebro sin caer en reduccionismos simplistas. Por otro, de la responsabilidad ética para garantizar que los sistemas artificiales amplíen, y no limiten, las capacidades humanas. Conceptos emergentes como los neuroderechos, la auditabilidad de modelos y el diseño seguro de sistemas híbridos serán esenciales para asegurar una relación equilibrada entre ambas inteligencias (Yuste et al., 2017).

En última instancia, pensar el cerebro junto con la inteligencia artificial no transforma únicamente nuestra forma de diseñar tecnología. También transforma la manera en que entendemos lo que significa aprender, recordar, olvidar y adaptarnos. Entre sinapsis que se fortalecen y algoritmos que se optimizan, surge un terreno conceptual donde las preguntas más antiguas de la filosofía —sobre la mente, el conocimiento y la conciencia— encuentran nuevas formas de expresarse. Allí, en el encuentro entre cables y sinapsis, entre señales eléctricas y datos, se dibuja el contorno de una inteligencia que es, al mismo tiempo, humana y artificial.

Referencias

Bassett, D. S., & Sporns, O. (2017). *Network neuroscience*. Nature Neuroscience, 20(3), 353–364. https://doi.org/10.1038/nn.4502

Bliss, T. V. P., & Lømo, T. (1973). *Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path*. Journal of Physiology, 232(2), 331–356.

Cao, Y., & Yang, J. (2015). *Towards making systems forget with machine unlearning*. In 2015 IEEE Symposium on Security and Privacy (pp. 463–480). https://doi.org/10.1109/SP.2015.35

Davies, M., Srinivasa, N., Lin, T. H., Chinya, G., et al. (2018). *Loihi: A neuromorphic manycore processor with on-chip learning*. IEEE Micro, 38(1), 82–99.

Edelman, G. M. (1987). *Neural Darwinism: The theory of neuronal group selection*. Basic Books.

Edlow, B. L., & Menon, D. K. (2024). *Brain network connectivity and consciousness*. Critical Care Medicine, 52(9), 1414–1426.

Faisal, A. A., Selen, L. P., & Wolpert, D. M. (2008). N*oise in the nervous*

system. Nature Reviews Neuroscience, 9(4), 292–303.

Fox, M. D. (2018). *Mapping human brain networks*. New England Journal of Medicine, 379(23), 2237–2245.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Isaacson, J. S., & Scanziani, M. (2011). *How inhibition shapes cortical activity*. Neuron, 72(2), 231–243.

Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., & Hudspeth, A. J. (2013). *Principles of neural science* (5th ed.). McGraw-Hill.

Lake, B. M., Ullman, T., Tenenbaum, J. B., & Gershman, S. (2017). *Building machines that learn and think like people*. Behavioral and Brain Sciences, 40, e253.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep learning*. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

McCulloch, W. S., & Pitts, W. (1943). *A logical calculus of the ideas immanent in nervous activity*. Bulletin of Mathematical Biophysics, 5, 115–133. https://doi.org/10.1007/BF02478259

Mountcastle, V. (1997). *The columnar organization of the neocortex*. Brain, 120(4), 701–722.

Nayak, M., et al. (2022). *Regional neuronal plasticity*. Heliyon, 8(12), e12292.

Rosenblatt, F. (1958). *The perceptron: A probabilistic model for information storage and organization in the brain*. Psychological Review, 65(6), 386–408. https://doi.org/10.1037/h0042519

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). *Learning representations by back-propagating errors*. Nature, 323(6088), 533–536.

Serruya, M. D., Hatsopoulos, N. G., Paninski, L., Fellows, M. R., & Donoghue, J. P. (2002). *Instant neural control of a movement signal*. Nature, 416(6877), 141–142.

Shenoy, K. V., & Carmena, J. M. (2014). *Combining decoder design and neural adaptation in brain–machine interfaces*. Neuron, 84(4), 665–680.

Sporns, O. (2011). *Networks of the brain*. MIT Press.

Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). *Attention is all you need. In Advances in Neural Information Processing Systems* (Vol. 30).

Xu, X., Ren, K., Yu, Z., & Zhang, S. (2023). *Machine unlearning: A survey. ACM Computing Surveys*, 55(8), 1–37. https://doi.org/10.1145/3527448

Yuste, R., et al. (2017). *Four ethical priorities for neurotechnologies and AI. Nature*, 551(7679), 159–163.

Joshsua J. González Díaz *es Ingeniero de Sistemas de la Pontificia Universidad Javeriana, Especialista en Seguridad de la Información de la Universidad de los Andes, Especialitsta en Derecho Informático y de las Nuevas Tecnologías de la Universidad Externado de Colombia, en conjunto con la Universidad Complutense de Madrid. Es Magíster en Seguridad de la Información de la Universidad de los Andes y Magíster en Derecho Informático y de las Nuevas Tecnologías del Externado. Se ha formado también como Gestor de Ciberseguridad para Gerentes en el MIT Sloan School of Management y actualmente es candidato a Doctor (PhD(c)) en Cybersecurity Analytics and AI en la George Washington University Se desempeña como Chief Information Security Officer (CISO), Data Privacy Officer (DPO) y Chief Compliance Officer (CCO) en una empresa internacional, liderando la estrategia de ciberseguridad, cumplimiento normativo y privacidad de datos. Cuenta con certificaciones internacionales como CEH, CHFI, ECSA, LPT, ISO/IEC 27001:2022 Lead Auditor, ISO 27032:2018 Cybersecurity Lead Manager, ISO 31000:2018 Risk Management e Incident Response and Forensics (CIRF), entre otras, que respaldan su amplia experiencia en gobierno, gestión y respuesta a incidentes de seguridad de la información.*

“Text to Anything”

La Inteligencia Artificial Generativa en entornos empresariales

DOI: 10.29236/sistemas.n177a8

Resumen

La IA generativa es considerada una tecnología de propósito general y se ha convertido en tema recurrente y central en el día a día de los individuos y las organizaciones. La abundante oferta de servicios, la rápida evolución de capacidades de estas herramientas y la feroz competencia entre los gigantes tecnológicos han despertado grandes expectativas entre los inversionistas, usuarios y observadores de estas tecnologías. Este documento da cuenta de un ejercicio de titulado “*Text to Anything*”, llevado a cabo en un ambiente universitario (nivel pregrado y posgrado) en ciencias empresariales. Se exploraron de manera guiada diversas herramientas y casos de uso de la inteligencia artificial para la creación de diferentes tipos de contenidos (texto, código, audio, video) a partir de indicaciones en lenguaje natural, mejor conocidos por el anglicismo: *prompts*. Este ejercicio se llevó a cabo en diferentes momentos y dejo entre otras las siguientes observaciones: la *primera* es que la velocidad de las innovaciones nos debe llevar crear estrategias de transformación digital organizacional. *Segundo*, es necesario profundizar la alfabetización digital de nuestros colaboradores para que habilite y desbloquee el potencial de la IA Generativa. *Tercero*, existe un desconocimiento generalizado del *cómo funcionan* estos sistemas. Por último, encontramos desconocimiento de las vulnerabilidades y los riesgos asociado al uso inadecuado de estas herramientas. Este artículo comenta sobre estos retos y provee una serie de recomendaciones para que las organizaciones contemporáneas puedan hacer una transición hacia la creación de valor a través de la experimentación ágil y la identificación de casos de uso escalables.

Palabras clave

Inteligencia artificial generativa, organizaciones, estrategia, experimentación

Introducción

La Inteligencia Artificial Generativa en sus diferentes capacidades (generación de texto, código de computador, imágenes, audio y video), tipos (*open source* o *proprietary*) y servicios más reconocidos (i.e. ChatGPT, Gemini, Claude, Grok, entre otros) es producto de la interacción de tecnologías de información y telecomunicaciones desarrolladas y perfeccionadas en las últimas décadas.

La IA Generativa es literalmente el resultado de un *stack* de infraestructura física y digital como los servicios en la nube, el internet, las redes sociales, los teléfonos inteligentes, dispositivos IoT (*Internet of Things*), gigantescas granjas de servidores, la manufactura de circuitos electrónicos, todo esto conectado globalmente por cableado submarino de alta velocidad, pero paradójicamente estos servicios son prestados a un usuario final vía inalámbrica, por lo tanto invisibilizando la compleja infraestructura que soporta estos servicios.

La inteligencia artificial ha sido conceptualizada en la literatura reciente como una tecnología de propósito general (Eloundou et al., 2024), pero a diferencia de otras innovaciones consideradas tecnologías de propósito general como la electricidad, el motor de combustión, los semiconductores y otras, los servicios de IA generativa han tenido una difusión y adopción global bastante veloz.

Adicionalmente, las barreras de entrada para *usar* estos servicios son relativamente bajas, para comenzar a usar alguno de los servicios de inteligencia artificial multimodal lo único que es necesario es tener disponible un dispositivo electrónico con una conexión a internet, el acceso a estos servicios está literalmente al alcance de la mano. Por otro lado, crear y proveer estos servicios tiene importantes necesidades en términos de infraestructura y energía, acceso a datos de calidad, y talento del más alto nivel de desempeño y esta oferta tecnológica está concentrada particularmente en China, Estados Unidos y la Unión Europea.

No se puede concebir ningún proceso productivo contemporáneo que no incorpore el uso de computadores y nuevas tecnologías en el diseño, producción, comercialización de bienes y servicios, es por eso por lo que la promesa de la inteligencia artificial se presenta como una oportunidad que pueden aprovechar las organizaciones en casi todos los sectores, pero de la misma forma, las organizaciones no han cerrado brechas importantes en la alfabetización digital de los colaboradores y la transformación digital de sus procesos.

El lanzamiento al mercado de servicios de Grandes Modelos de Lenguaje (LLM, por sus siglas en inglés) como ChatGPT, Claude y Gemini ha transformado de manera radical el panorama organizacional, permitiendo el análisis de vastos conjuntos de datos textuales con una rapidez y eficiencia sin precedentes. Debido a su capacidad para procesar y comprender el lenguaje natural, los LLMs están redefiniendo como las organizaciones extraen y procesan información valiosa, identifican patrones y generan conocimientos estratégicos a partir de datos previamente inaccesibles o subutilizados. Su impacto se extiende a diversas áreas como la toma de decisiones basadas en datos, la automatización de procesos rutinarios y/o complejos, la personalización de servicios entre otros.

El poder de la palabra: *From Text to Anything*

La oferta tecnológica de inteligencia artificial generativa es muy amplia. En nuestro ejercicio, estuvimos experimentando con los diversos modelos de inteligencia artificial disponibles de manera *freemium,* es decir, acceso básico gratis y funcionalidades avanzadas bajo modelo de suscripción. Estos servicios, se encuentran ranqueados en el sitio LM Arena Leaderboard <https://lmarena.ai/leaderboard>, y dado que los modelos, sus capacidades y versiones se cambian, se actualizan, se robustecen y/o se hacen obsoletos es clave invitar a una experimentación abierta a las diferentes herramientas disponibles.

Existen aplicativos de IA generativa para casi todo. Estamos ante un escenario de ***text-to-anything*** o la creación de todo tipo de contenido digital a partir de indicaciones de lenguaje natural, conocidos en el contexto de la inteligencia artificial como *prompts*. Los casos de uso más frecuentes y conocidos son la creación de documentos y presentaciones, creación de contenidos audio visuales, diseño asistido por computador (CAD, *Computer Assisted Design*), generación de código de computador para crear aplicaciones o videojuegos. Literalmente, el límite es la creatividad y las habilidades del humano que escribe el *prompt*, su conocimiento tácito, sus sensibilidades estéticas, en resumen, la inteligencia artificial amplifica las potencialidades del trabajo del conocimiento —(Jarrahi et al., 2023; Pai et al., 2022).

La clave está en el *prompt*

Los sistemas de inteligencia artificial generativa contemporáneos poseen la capacidad de optimizar las instrucciones que reciben a través de la formulación de preguntas contextualizantes al usuario. Así mismo, se pueden utilizar las herramientas para perfeccionar un *prompt*, es decir crear un *meta-prompt*. Un *meta-prompt* se define como una instrucción de segundo orden diseñada para generar o perfeccionar otras instrucciones. En

Tabla 1
Text to Anything: Capacidades actuales de la IA generativa

Capacidades	Ejemplos
Text to Text Generación automatizada de contenido textual estructurado mediante procesamiento de lenguaje natural avanzado	Redacción de documentos académicos, correspondencia formal, informes técnicos, presentaciones ejecutivas y material didáctico, todos producidos mediante instrucciones expresadas en lenguaje natural.
Text to Code Síntesis de código fuente funcional en múltiples paradigmas de programación	Desarrollo de fragmentos de código (*snippets*) en diversos lenguajes de programación, construcción de interfaces web responsivas, implementación de aplicaciones móviles y de escritorio, mediante instrucciones en lenguaje natural (*prompts*).
Text to Image Generación de representaciones visuales mediante modelos de difusión y arquitecturas neuronales profundas	Creación de ilustraciones, fotografías sintéticas, diseños gráficos y contenido visual artístico a partir de descripciones textuales detalladas.
Text to Sound Producción de contenido sonoro y musical mediante redes neuronales especializadas en síntesis acústica.	Composición musical automatizada, diseño de efectos sonoros, síntesis de voz y creación de paisajes sonoros complejos.
Text to Video Generación de secuencias audiovisuales dinámicas mediante la integración de modelos de difusión temporal y síntesis multimodal	Producción de contenido cinematográfico, material audiovisual, animaciones, narrativas y presentaciones multimedia sincronizadas.
Text to Game Creación de entornos ludificados y experiencias de juego mediante motores de desarrollo asistidos por IA.	Prototipado rápido de videojuegos, diseño de mecánicas de juego y desarrollo de experiencias interactivas inmersivas.

Fuente: Elaboración propia

esencia, constituye un "*prompt* para crear *prompts*", permitiendo la transformación de indicaciones rudimentarias o ambiguas en especificaciones robustas, exhaustivas y técnicamente precisas.

La importancia de un buen *prompt* se puede ilustrar mejor en el aforismo: "*Pedís y no recibís porque pedís mal.*" Este breve adagio en este contexto de la IA nos enseña que para interactuar de manera óptima

con la IA generativa es necesario que escribir con claridad las indicaciones deseadas. La estrategia iterativa de nuestro ejercicio fue replicar, reproducir o hacer derivaciones de listados de *prompts* encontrados en internet. Algunos generados desde los mismos proveedores de estos servicios para dar a conocer las funcionalidades de sus herramientas. En la tabla 2 algunas de las fuentes que usamos para guiar los *prompts* de nuestros ejercicios.

El imperativo organizacional es entonces una actualización de las operaciones ante la proliferación de datos, y la democratización de herramientas de aprendizaje automático (Machine Learning, en inglés) e inteligencia artificial que reducen la necesidad de un conocimiento extenso de programación gracias a soluciones de bajo código (low-code) y sin código (no-code), exige una revisión de los enfoques gerenciales actuales y será un llamado a la formación en nuevos alfabetismos y en nuevas buenas prácticas en un entorno altamente digitalizado, con gran abundancia en datos pero con grandes retos en términos de analítica (Jawad & Balázs, 2024).

Según diversas estimaciones, 80% de los datos de las organizaciones son datos no estructurados, es decir no tabulares y usualmente incluyen, texto, imágenes y video (Drummer, 2020), por lo estos tipo de datos implican una dificultad para tratar con funciones de analítica tradicionales, sin embargo, los grandes modelos de lenguaje multimodales (LLM) han abierto una oportunidad enorme para que las organizaciones capitalicen sobre el valor que tienen los datos no estructurados alojados en servidores, en nubes corporativas y otros tipos de almacenamiento.

Retos inmediatos para las organizaciones

Es claro que la inteligencia artificial es una tecnología de propósito general que tiene el potencial de transformar profundamente las organizaciones modernas, sin embargo, las organizaciones necesitan conocer su estado actual de digitalización a través de una medición interna de su madurez digital. Estos modelos de madurez no solo incorporan variables e indicadores relativamente comparables, sino que también deben dar cuenta del diseño único de los procesos y la cultura organizacional dominante en la organización lo que implica una ardua tarea de reflexión por parte del liderazgo corporativo.

Por otro lado, es necesario discutir los “nuevos alfabetismos” y competencias digitales necesarios para alcanzar el máximo potencial de estas tecnologías en las organizaciones. Los alfabetismos digitales, competencias y habilidades digitales han sido estudiadas ampliamente en la literatura (Tinmaz et al., 2022), sin embargo, aún se presentan grandes brechas de difu-

Tabla 2
Repositorio Breve de Prompts

Fuente	Enlace	QR
Claude Prompt Library	https://docs.claude.com/en/resources/prompt-library/library	
ChatGPT Prompt Packs	https://academy.openai.com/public/tags/prompt-packs-6849a0f98c613939acef841c	
Gemini Prompts	https://ai.google.dev/gemini-api/prompts?hl=es-419	
Midjourney Explore	https://www.midjourney.com/explore	
Prompts.chat	https://prompts.chat/	
Superhuman 1000+ Prompts	https://academy.superhuman.ai/c/1-000-prompts	

Fuente: Elaboración propia

sión y adopción en las organizaciones, en parte por el acelerado proceso de innovación en términos de digitalización e inteligencia artificial lo que implica una gran velocidad en la obsolescencia del conocimiento adquirido por los individuos.

La correcta adopción de aplicaciones de inteligencia artificial en las organizaciones requiere una comprensión profunda no solo de qué son y cómo operan los modelos de lenguaje (Wolfram, 2023), sino también de los principios técnicos subyacentes, como el mecanismo de atención, introducido por los *Transformers* (Vaswani et al., 2017). Este mecanismo es fundamental, ya que permite a los modelos procesar grandes cantidades de datos textuales y enfocarse en las partes más relevantes de la información, optimizando así la generación de respuestas contextualmente adecuadas.

Además, entender estas bases técnicas permite a las organizaciones evaluar con mayor precisión las capacidades y limitaciones de los sistemas de Inteligencia Artificial, alineando su implementación con objetivos estratégicos claros. Esto no solo garantiza un uso más eficiente y ético de las tecnologías, sino que también fomenta una adopción más integrada en los procesos internos, desde la toma de decisiones hasta la mejora en la experiencia del cliente. En resumen, una adopción informada facilita la maximización del valor generado por estas tecnologías, al tiempo que minimiza riesgos asociados, como interpretaciones erróneas o sesgos no deseados en los resultados. En suma, se hace fundamental reconocer que los LLMs no son oráculos que tienen todas las respuestas correctas, sino que tienen fuertes sesgos dados los datos de entrenamiento, valores y los *prompts* de sistema impuestos por los desarrolladores de cada modelo.

Por último, se hace urgente discutir los retos en términos de la seguridad digital que presentan los modelos de lenguaje. Infiltraciones y exfiltraciones de datos sensibles, privados, confidenciales o secretos, la evidencia anecdótica encontrada en medios de comunicación incluye el caso del gigante de la electrónica Samsung prohibiendo el uso de chatbots luego de la filtración de información sensible de la empresa (Gurman, 2023). También se evidencia en el caso Moffatt v. Air Canada en el que un chatbot dio información incorrecta a un usuario y la aerolínea fue obligada por una corte a pagar daños y perjuicios '"(Cecco, 2024). En el siguiente enlace una revisión exhaustiva de riesgos de seguridad asociados al uso de IA del MIT AI Risk Repository. < https://airisk.mit.edu/ >

Hacia la creación y la captura de valor en las organizaciones contemporáneas

En conclusión, la inteligencia artificial se ha consolidado como un ele-

mento clave en la transformación de las organizaciones, ofreciendo oportunidades sin precedentes para generar valor y competitividad.

Sin embargo, para que estas oportunidades se materialicen, es fundamental superar el entusiasmo inicial (*hype*) que rodea a esta tecnología y avanzar hacia su integración estratégica y práctica. Este proceso requiere una comprensión profunda de cómo las tecnologías de Inteligencia Artificial pueden alinearse con los objetivos organizacionales, asegurando que sus aplicaciones no sean simplemente innovadoras, sino también funcionales y relevantes para las metas a largo plazo.

Un punto central en esta transición es la necesidad de adoptar una mentalidad digital que permita entender tanto el potencial como las limitaciones de la inteligencia artificial. Esto incluye realizar un análisis crítico del estado actual de la organización en términos de su transformación digital (AS IS) y establecer una visión clara del estado deseado (TO BE). Este enfoque estratégico no solo permite identificar las brechas existentes, sino también priorizar las áreas clave para la implementación de IA, asegurando que los esfuerzos estén alineados con las necesidades reales de la organización y su entorno competitivo.

Además, la experimentación desempeña un papel crucial en el proceso de adopción de inteligencia artificial. Probar y evaluar casos de uso específicos permite a las organizaciones validar hipótesis, identificar oportunidades de mejora y mitigar riesgos antes de realizar inversiones a gran escala. Este enfoque iterativo fomenta una cultura organizacional orientada a la innovación, donde los equipos tienen la flexibilidad y el respaldo necesarios para aprender de los errores y refinar sus estrategias.

La experimentación, por lo tanto, no solo acelera la curva de aprendizaje, sino que también contribuye a una implementación más efectiva y adaptada a las particularidades de cada organización.

Finalmente, para garantizar el éxito de las iniciativas de inteligencia artificial, es esencial contar con un inventario sólido de datos y un enfoque centrado en el desarrollo de habilidades clave en los colaboradores. Los datos deben ser de alta calidad, accesibles y gobernados bajo principios éticos y normativos, mientras que los equipos deben estar capacitados para interpretar, gestionar y maximizar el valor que se puede extraer de estas tecnologías.

Solo a través de un enfoque inte-gral que combine estrategia, experimentación, gestión de datos y desarrollo de talento, las organizaciones podrán convertir la inteligencia artificial en un motor de transformación sostenible y tangible.

Referencias

Cecco, L. (2024, February 16). *Air Canada ordered to pay customer who was misled by airline's chatbot.* The Guardian. https://www.theguardian.com/world/2024/feb/16/air-canada-chatbot-lawsuit

Drummer, A. (2020, November 19). *Extracting insights from complex, unstructured big data* | IBM. https://www.ibm.com/think/insights/managing-unstructured-data

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2024). *GPTs are GPTs: Labor market impact potential of LLMs*. Science, 384(6702), 1306–1308. https://doi.org/10.1126/science.adj0998

Gurman, M. (2023, May 1). *Samsung Bans ChatGPT, Google Bard, Other Generative AI Use by Staff After Leak—Bloomberg*. Bloomberg. https://www.bloomberg.com/news/articles/2023-05-02/samsung-bans-chatgpt-and-other-generative-ai-use-by-staff-after-leak

Jarrahi, M. H., Askay, D., Eshraghi, A., & Smith, P. (2023). *Artificial intelligence and knowledge management: A partnership between human and AI*. Business Horizons, 66(1), 87–99. https://doi.org/10.1016/j.bushor.2022.03.002

Jawad, Z. N., & Balázs, V. (2024). *Machine learning-driven optimization of enterprise resource planning (ERP) systems: A comprehensive review*. Beni-Suef University Journal of Basic and Applied Sciences, 13(1), 4. https://doi.org/10.1186/s43088-023-00460-y

Pai, R. Y., Shetty, A., Shetty, A. D., Bhandary, R., Shetty, J., Nayak, S., Dinesh, T. K., & D'souza, K. J. (2022). *Integrating artificial intelligence for knowledge management systems – synergy among people and technology: A systematic review of the evidence*. Economic Research-Ekonomska Istraživanja, 35(1), 7043–7065. https://doi.org/10.1080/1331677X.2022.2058976

Tinmaz, H., Lee, Y.-T., Fanea-Ivanovici, M., & Baber, H. (2022). *A systematic review on digital literacy*. Smart Learning Environments, 9(1), 21. https://doi.org/10.1186/s40561-022-00204-y

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need* (arXiv:1706.03762). arXiv. http://arxiv.org/abs/1706.03762

Wolfram, S. (2023). *What is ChatGPT doing ... And why does it work?* Wolfram Media, Inc.

Carolina Saldaña Cortes.
Profesional con más de veinte años de experiencia en consultoría y veinticinco en docencia e investigación universitaria, especializado en analítica de negocios y en el desarrollo de modelos predictivos y prescriptivos para la toma de decisiones estratégicas. Desde 2022 dirige el Centro de Tecnología y Analítica de la Universidad Externado de Colombia, donde coordina programas de posgrado y proyectos de consultoría orientados a la Transformación Digital y Analítica de Negocios. Director del Centro de Tecnologia y Analitica, Facultad de Administración de Empresas, Universidad Externado de Colombia (jenny.saldana@uexternado.edu.co)

Andrés Aguilera Castillo.
Consultor empresarial, profesor universitario y conferencista en áreas de Transformación Digital e Inteligencia Artificial para las organizaciones. Actualmente es profesor de tiempo completo en el Programa de Negocios Internacionales del Tecnológico de Monterrey, Campus Puebla. Profesor a Tiempo Completo en la Escuela de Negocios del Tecnológico de Monterrey Campus Puebla (andres.aguilera@tec.mx)

La burbuja de la IA generativa: “a feature not a bug”

DOI: 10.29236/sistemas.n177a9

Resumen
Este artículo se pregunta si hay una burbuja especulativa de inteligencia artificial generativa. Responde que sí, pero que esto es una simplificación que de poco nos sirve para tomar decisiones. Las burbujas son parte inevitable y necesaria de toda revolución tecnológica, tras su estallido se da paso a los nuevos modelos que se generan. Además, en entornos de innovación es natural que solo pocos sean exitosos. Ya antes se vio un fenómeno similar en la burbuja puntocom, pero ahora las burbujas pueden ser más rápidas, simultáneas o extendidas. En consecuencia, invertir mejor en IA implica tener una mirada del ecosistema digital y organizaciones ambidiestras, capaces de asumir riesgos de manera equilibrada para innovar exitosamente.

Palabras claves
Inteligencia artificial, burbuja especulativa, ecosistemas digitales, organizaciones ambidiestras

Introducción

Se discute cada vez en medios noticiosos y académicos la existencia de una burbuja, puede ser de la tecnología, o de la inteligencia artificial en particular. La valoración de las empresas de IA alcanza niveles exorbitantes, pero ya hay dudas sobre su verdadero valor para las empresas que la han empezado a usar o que han invertido grandes sumas en ella. MinCiencias acaba de lanzar y cerrar una convocatoria por 630 mil millones de pesos (MinCiencias, 2025), demostrando que es un fenómeno global y de primera prioridad. Naturalmente, esto genera incertidumbre en actores públicos y privados sobre si invertir, cuánto y cómo.

En este artículo, primero preguntamos si hay tal burbuja de IA. A ello respondemos que sí, aclarando qué es una burbuja y cuáles son las cifras que así lo indican. La siguiente sección cuestiona la utilidad de simplificar la respuesta, pues las burbujas son parte del modelo de crecimiento de nuevas tecnologías y necesarias para la generación del nuevo ordenamiento.

En la siguiente sección, se dan dos recomendaciones puntuales: los ecosistemas digitales y las organizaciones ambidiestras, como posibles lineamientos para mejores decisiones de inversión en IA.

¿Hay una burbuja de inteligencia artificial generativa?

Una burbuja de especulación financiera existe si la única razón por la que el precio de algo es alto hoy es que los inversionistas creen que el precio de venta será alto mañana y los factores fundamentales no justifican un precio así ——(Stiglitz, 1990). El precio aquí se refiere a los activos financieros (por ejemplo, los precios de las acciones de las empresas) y no de consumo, pues implica una posible reventa futura por parte de un inversionista. Nos podemos enfocar entonces en el valor de las empresas de inteligencia artificial (IA), no en el precio de sus servicios. Ahora, particularmente con la IA generativa, muchas de las empresas son aún emergentes y privadas, por ende, no se puede hablar de precio de acciones, aunque sí de una valoración.

Entre las privadas, la valoración de OpenAI llegó a los quinientos mil millones de dólares este octubre Hu, 2025) y la de Anthropic a trescientos cincuenta mil millones este noviembre (Seiler, 2025). Esta última, por cierto, estaba valorada en la mitad un mes antes y en un quinto seis meses más atrás (Reuters, 2025), demostrando no solo valores extraordinarios sino exponencialmente crecientes. Hay otro conjunto de empresas ya establecidas

y públicas que, sin ser de IA, han crecido porque la IA requiere su infraestructura. Entre estas, Nvidia está valorada en cinco billones (se dice *trillions* en EE.UU.) de dólares (Nishant & Singh, 2025). Esta cifra, nunca antes vista, hace de Nvidia la empresa más valiosa del mundo y una de las razones principales por la que se habla de la burbuja de la IA hoy.

Lejos de anticiparlo, muchos esperaban que este año Nvidia se desinflara. Muchos pesimistas, claro. Por eso, para que haya una burbuja especulativa se requiere más que solo la diferencia entre el precio y el valor fundamental. Se requiere un mercado que permita tanto a pesimistas como optimistas "opinar" con su dinero: debe haber un volumen suficiente de transacciones (Porras, 2017). Esto claramente lo hay, no son solo OpenAI, Antropic o Nvidia, las gigantes como Amazon, Google y Microsoft también han crecido sobre la ola de la IA.

Decimos que para una burbuja, además de la diferencia entre precio y valor real, se requiere un alto volumen de transacciones y este, a su vez, depende de que haya suficiente acceso a capital (dinero barato, por ejemplo) (ibid 2017). En efecto, otra razón por la que algunos hablan de burbuja en tecnología y en IA en particular es porque la era del dinero barato se había acabado '(Daniel, 2022), pero al parecer ni SoftBank ni los fondos saudíes se han enterado, a juzgar por los ejemplos de arriba. En la post-pandemia, el desinfle de Zoom, los despidos en Amazon y Microsoft o la falta de resultados de Reality Labs de Meta se citan como muestras de la burbuja desinflándose. Meta, tras el cambio de su nombre y una inversión que ya acumula 46 mil millones de dólares, no logra sacarle provecho al metaverso al que apostó (Levy, 2024). Pero, aunque el metaverso no haya despegado, Meta sigue creciendo; de hecho, sobre la ola de la IA y, sobre todo, lo hace al reinvertir en sí misma.

Porque además de los fondos de riesgo, la inversión circular es otra manera de mantener el flujo de capital requerido para el crecimiento. Eso señala otro fenómeno que agrega complejidad a la discusión en torno a la burbuja. No solo Meta, sino Microsoft y Apple también son conocidas por su circularidad de flujos (*stock buyback*) que ya no solo aplican sobre si mismas sino en complejas cadenas y redes de inversión circular o bidireccional, como en el Project Stargate que incluye a Microsoft, Nvidia, OpenAI y Anthropic (OpenAI, 2025). En Stargate participa incluso Oracle, aunque de manera muy diferente, endeudándose al punto de preocupar a los bancos '(Mutua, 2025). En efecto, esa es otra fuente adicional de capital, el endeudamiento.

Por ello, el riesgo y la proyección futura no son iguales para todos los actores. Cuando se es líder de la revolución tecnológica a punta de

inversión de riesgo (OpenAI) es muy diferente que cuando se trata de la transformación de un negocio ya consolidado a punta de préstamos (Oracle). Mientras que una apuesta con el dinero de los demás y tiene máximo interés en que la especulación crezca para su nueva ronda de financiación, la otra asume mayor riesgo y no está interesada en especulación, sino en retornos para pagar la deuda. La preocupación de quienes apuestan sus propios recursos es natural. Si el efecto dominó hace que muchas crezcan, ¿hará también que muchas caigan?

Conviene acordarnos de Groupon, empresa de tecnología que en su momento fue la de más rápido crecimiento de la historia. El día que salió a bolsa hace catorce años tocó su máxima valoración de 18.8 mil millones de dólares y los pesimistas anunciaron con razón la burbuja que ese mismo día se desinfló notablemente y desde entonces no dejó de hacerlo; hoy sobrevive con menos del 1% de su máximo valor (Lunden, 2023). Burbuja, sí, para los inversionistas de Groupon, pero es que su exitoso modelo de negocio era fácilmente imitable y las compras grupales en línea simplemente se integraron al comercio electrónico. Otras empresas no desparecieron por el hecho de que la burbuja Groupon se desinflara; no hubo efecto dominó que las arrastrara hacia abajo. Más bien fue al contrario, emergieron nuevas compañías y servicios en las ya existentes que le quitaron la mayor parte del mercado a Groupon.

Algo similar se señala recurrentemente cuando nos remitimos a la burbuja puntocom de finales del siglo pasado y se resalta que, pese al descalabro, de allí salieron justamente los gigantes que siguen ahora en el medio de la burbuja actual: Google y Amazon, entre otros. Entonces, ni el estallido de la burbuja de una empresa notablemente grande o acelerada implica el fin de todo el sector que lidera, ni tampoco el desinflarse de un sector implica el desinflarse de todas sus empresas.

"It's not a bug, it's a feature"

En tecnología deberíamos estar acostumbrados a esta dinámica, la valoración de Facebook, Google o Twitter fue alta incluso mientras generaban pérdidas, porque más adelante vendrían las ganancias. Pueden ser pocos o muchos años para que eso ocurra y la habilidad de los optimistas y emprendedores es empujar el horizonte cada vez más al futuro. Sam Altman y otros líderes de la industria de IA han ido corriendo la fecha en que esperan que llegue la IA general. En el camino también van ajustando expectativas sobre lo que es esa inteligencia general (o superinteligencia, o tal vez solo IA agencial); hoy anuncian la singularidad en cinco años '(Quiroz-Gutierrez, 2025). He aquí el diseño de la burbuja: por un lado, con la ley de escalamiento (*scaling law*) anuncian la inevitable

llegada de la IA general; pero, por otro lado, ese escalamiento no es tan inevitable, pues depende de la capacidad, es decir de que la inversión siga creciendo.

Pero no es solo la IA, llevamos años impulsando la transformación digital con otras tecnologías de la cuarta revolución industrial, bajo la creencia de que son radicalmente transformadoras y requerirán de inversiones grandes. Cuando llegan esas inversiones no deberíamos sorprendernos. Incluso si no se trata de una burbuja "diseñada", es un patrón que todos reconocemos. La burbuja es una fase estructural de cada revolución tecnológica: el capital financiero impulsa una expansión especulativa que infla el valor de las nuevas tecnologías y de las industrias asociadas más allá de su capacidad productiva real. Esa desconexión genera tensiones acumuladas que culminan en un estallido, cuyo colapso abre el espacio político e institucional para corregir excesos, redefinir reglas y reorientar el financiamiento hacia usos productivos (Perez, 2002). La burbuja no es un accidente ni un fenómeno aislado, sino el mecanismo mediante el cual el sistema rompe con la inercia del paradigma anterior, moviliza recursos para construir las nuevas infraestructuras y, tras la crisis, permite el despliegue ordenado del nuevo modelo de desarrollo.

Lo que ocurre es que con la creciente aceleración y complejidad, puede ser que una sola burbuja no baste para estabilizar y, como en el caso puntocom, requiera de una segunda burbuja (la crisis financiera de 2008) que es en realidad extensión de la primera ——"(Perez, 2009).

Así que la respuesta a nuestra pregunta inicial puede ser general y rotunda: sí hay una burbuja de IA generativa. Pero, ¿qué hacemos con esa respuesta en un contexto en que las burbujas no son accidentales, sino necesarias y además se pueden superponer o extender? Y ¿de qué sirve saber que hay una burbuja en IA generativa, si no estamos ante la revolución de una sola tecnología, sino de un conjunto de tecnologías convergentes y rápidamente cambiantes?

Cómo responder a la burbuja

Decir que hay una burbuja puede ser una simplificación extrema y conducirnos a desestimar el fenómeno simplemente como un riesgo del mercado. Ya pasamos por ahí: como consecuencia de la burbuja puntocom, el editor del Harvard Business Review puso en tela de juicio a todo el sector informático con su controversial "IT doesn´t matter" Carr, 2003). Allí señalaba la ruptura de la burbuja como un claro aviso de que la tecnología ya había llegado al fin de su despliegue y solo pocas compañías estarían excepcionalmente en capacidad de seguir sacando ventaja. ¿Las demás? Las invitaba a la retirada: gasten menos, sigan en vez de li-

derar y enfóquense en el riesgo, no en la oportunidad de la tecnología de información.

¿Qué tal que le hubiéramos hecho caso? Tal vez Carr no entendía que ya Robert Solow había señalado la paradoja de la productividad en la tecnología de información. Lo usual, por si no era claro, es que haya una diferencia estadística entre la inversión en TI y los aumentos en productividad. Pero lo usual es que esto se deba a no entender que el determinismo tecnológico no existe, que el efecto de la inversión toma tiempo, que no medimos la productividad donde o cuando se debe, que las expectativas son exageradas o que simplemente no se gestiona adecuadamente la inversión —'(Schweikl & Obermaier, 2020). Lo usual es que, en contextos de innovación, por naturaleza, solo a unos pocos les va bien, a la mayoría no.

Así que no se trata de no invertir sino de invertir bien. A sabiendas de que la probabilidad de éxito se reserva a lo mejor. Dos abordajes que podrían orientar estas inversiones son los ecosistemas digitales y las organizaciones ambidiestras. Al tomar una perspectiva ecosistémica se entiende que no hay tal cosa como IA aislada, sino una red de tecnologías, recursos minerales, fuentes de energía, proveedores, academia, servicios, sistemas de pago, talento, regulaciones, expectativas, emociones y tendencias geopolíticas que deben ser parte del diseño de la inversión, porque sin ellas no funciona la IA. No basta invertir en infraestructura, en datos, en capacitación, en diseño de nuevos productos o servicios por separado y esperar retornos. Y estarán mejor posicionadas para navegar esa complejidad de alto riesgo las organizaciones ambidiestras, capaces de equilibrar sus esfuerzos de exploración (nuevos desarrollos en IA) con sus actividades de explotación (seguir operando y vendiendo cada vez mejor lo que ya hacen).

Conclusiones

Sí hay una burbuja de inteligencia artificial generativa, pero no debería ser sorpresa, porque las burbujas son parte intrínseca de toda revolución tecnológica. Más bien, deberíamos asumir el crecimiento de inversiones como parte del ecosistema de innovación digital e invertir mejor los recursos con una mirada más compleja y ambidiestra.

Referencias

Carr, N. G. (2003, mayo). IT Doesn't Matter. Harvard Business Review. https://hbr.org/2003/05/it-doesnt-matter

Daniel, W. (2022, diciembre 28). 2022 killed the cheap money era. Here's what the next decade has in store. Yahoo Finance. https://finance.yahoo.com/news/2022-killed-cheap-money-era-130000862.html

Hu, K. (2025, octubre 2). OpenAI hits $500 billion valuation after share sale to SoftBank, others, source says | Reuters. Reuters.

https://www.reuters.com/technology/openai-hits-500-billion-valuation-after-share-sale-source-says-2025-10-02/

Levy, A. (2024, abril). Meta Platforms Has Spent $46 Billion on the Metaverse Since 2021, But It's Spending Twice As Much on This 1 Thing | Nasdaq. https://www.nasdaq.com/articles/meta-platforms-has-spent-$46-billion-on-the-metaverse-since-2021-but-its-spending-twice-as

Lunden, I. (2023, marzo 31). Groupon, which has lost 99.4% of its value since its IPO, names a new CEO... Based in Czech Republic. TechCrunch. https://techcrunch.com/2023/03/31/groupon-which-has-lost-99-4-of-its-value-since-its-ipo-names-a-new-ceo-based-in-czech-republic/

MinCiencias. (2025, octubre). Convocatoria Colombia Inteligente: Infraestructura Para el Desarrollo de la Inteligencia Artificial. Minciencias. https://minciencias.gov.co/convocatorias/plan-convocatorias-asctei-2025-2026/convocatoria-colombia-inteligente-infraestructura

Mutua, C. x. (2025, noviembre 27). Morgan Stanley Warns Oracle Credit Protection Nearing Record High. Yahoo Finance. https://finance.yahoo.com/news/morgan-stanley-warns-oracle-credit-215223845.html

Nishant, N., & Singh, R. (2025, octubre 29). Nvidia hits $5 trillion valuation as AI boom powers meteoric rise. Reuters. https://www.reuters.com/business/nvidia-poised-record-5-trillion-market-valuation-2025-10-29/

OpenAI. (2025, enero). Announcing The Stargate Project. https://openai.com/index/announcing-the-stargate-project/

Perez, C. (2002). Technological Revolutions and Financial Capital. https://www.elgaronline.com/monobook/9781840649222.xml

Perez, C. (2009). The double bubble at the turn of the century: Technological roots and structural implications. Cambridge Journal of Economics, 33(4), 779-805. https://doi.org/10.1093/cje/bep028

Porras, E. R. (2017). Bubbles and Contagion in Financial Markets, Volume 2. Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-52442-3

Quiroz-Gutierrez, M. (2025, septiembre). Sam Altman thinks AI will surpass human intelligence by 2030. His rival AI billionaires say it'll be even sooner | Fortune. Fortune. https://fortune.com/2025/09/26/sam-altman-openai-ceo-superintelligence-technology/

Reuters. (2025, septiembre 2). Anthropic's valuation more than doubles to $183 billion after $13 billion fundraise. Reuters. https://www.reuters.com/business/anthropics-valuation-more-than-doubles-183-billion-after-13-billion-fundraise-2025-09-02/

Schweikl, S., & Obermaier, R. (2020). Lessons from three decades of IT productivity research: Towards a better understanding of IT-induced productivity effects. Management Review Quarterly, 70(4), 461-507. https://doi.org/10.1007/s11301-019-00173-6

Seiler, G. (2025, noviembre 24). Microsoft and Nvidia Just Signed a Multibillion-Dollar Deal With Anthropic.

Here's What It Really Means for Investors. Yahoo Finance. https://www.fool.com/investing/2025/11/24/microsoft-and-nvidia-just-signed-a-multibillion-do/

Stiglitz, J. E. (1990). Symposium on Bubbles. Journal of Economic Perspectives, 4(2), 13-18. https://doi.org/10.1257/jep.4.2.13

Rafael A. González. *Ingeniero de Sistemas de la Pontificia Universidad Javeriana, Magíster en Ciencias de la Computación y Doctor en Ingeniería de Sistemas (cum laude) por la Universidad de Delft, en los Países Bajos. Como profesor titular de la Javeriana se desempeña investigador y consultor en sistemas de información, gestión del conocimiento y analítica de datos. Actualmente es editor de la revista Engineering for Development.*

Los mitos y realidades de la IA

DOI: 10.29236/sistemas.n177a10

Resumen

En este artículo se presentan al lector no técnico, algunos de los mitos actuales alrededor de la Inteligencia Artificial (IA) incluyendo conceptos para aclararlos. Novedad, crisis existenciales que plantea a los no legos, y otros. El objetivo principal es entender las novedades tecnológicas y evaluar así su mejor utilización. Posteriormente el autor se concentra en cuatro temas importantes para el futuro de la IA y de la computación avanzada: la metadata de los objetos generados computacionalmente por la IA, la importancia del contexto en el entendimiento del lenguaje natural, la regulación al desarrollo y al uso de la IA, para concluir en la importancia de la organización inteligente.

Palabras claves

Inteligencia artificial, mitos, metadata, aprendizaje automático.

Introducción

A manera de introducción, dejo al lector estas inquietudes: ¿Qué es la inteligencia? ¿Cuándo calificamos a una máquina de tener un comportamiento inteligente?

Los mitos alrededor de la Inteligencia Artificial

A continuación, presento los mitos más comunes asociados a la IA y a su utilización en la sociedad moderna. Importante resaltar la realidad asociada y así entender mejor la novedad tecnológica.

1. "La IA es nueva"

La IA incluye los campos de teoría de juegos, análisis de imágenes, robótica, análisis de lenguaje natural y otros similares, tratando de dotar a los computadores y máquinas derivadas, con capacidades humanas. Podemos decir que el objetivo de esta área de la computación es la máquina antropomorfa con todas las capacidades de interacción y razonamiento humano.

La investigación en análisis del lenguaje natural, el que hablan las personas, data de 1933 (Le Scao, 2020) y a pesar de estos más de noventa años de investigación y desarrollos, todavía estamos lejos de la traducción totalmente automática. Ambigüedades semánticas, regionalismos, juegos de palabras, seguirán siendo retos para los investigadores y desarrollos en esta área.

De manera similar la robótica (Thompson, 2021) en 1959 introducía la primera máquina en una planta de General Motors, y en 1972 producía máquinas que "observaban" su entorno, pudiendo crear un plan y ejecutarlo, precediendo las aspiradoras que desde 2002 recorren un apartamento evitando obstáculos.

En resolución de problemas, superando la capacidad humana, en 1997 (IBM, s.f.) el computador denominado "Deep Blue" venció en un juego de ajedrez al campeón mundial Gary Kasparov, marcando un hito en nuestra historia. "Deep Blue fue capaz de evaluar 200 millones de posiciones de ajedrez por segundo, logrando una velocidad de procesamiento de 11,38 mil millones de operaciones de punto flotante por segundo, o flops" (IBM, s.f.). En paralelo se desarrollaron los "sistemas expertos" en los 70s en el ambiente universitario (Feigenbaum, 1992), (Harmon 1985), dando lugar a prototipos de ayuda al diagnóstico médico (MYCIN), a la configuración de computadores (XCON), análisis de datos de espectrómetro de masas para identificar compuestos químicos (DENDRAL), entre otros. Estos prototipos y la investigación en sistemas expertos, realmente se vio limitada por la capacidad de los computadores de los 90s y no dieron lugar a productos formales. Hoy día se de-

sarrollan sistemas de propósito específico (similares a los mencionados), basados en reglas y extendiéndolas con capacidades de aprendizaje automático (Preis, 2023).

2. “El hombre ya no es el ser más inteligente ni el más capaz”

Las capacidades humanas han sido sobrepasadas por las máquinas desde la primera de sus invenciones: la palanca. Nacemos con máquinas conocidas y naturales a nuestro entorno, por lo que nos sorprende encontrar ahora computadores con capacidades de deducción, pero el tener habilidades matemáticas y poder procesar grandes volúmenes de información, en tiempos muy cortos, ya son habituales para nosotros. Podemos afirmar que el hombre nunca ha sido el ser más fuerte, ni el más capaz, y (desde que tenemos máquinas en nuestro apoyo) tampoco es el más hábil con los números o el más inteligente.

3. “La IA tiene conciencia y será autónoma”

El test de Turing, inicialmente llamado por Alan Turing como “el juego de la imitación”, consiste en que una persona interactúa aleatoriamente con otra persona y con un computador, a través de barreras y utilizando un “teleprinter”. Si el interlocutor principal no puede diferenciar cuando recibe mensajes del computador y cuando de la persona, el programa de computador utilizada era clasificado de “inteligencia artificial”. En palabras de Alan Turing (Turing, 1950) la duda *"¿Pueden pensar las máquinas?" debería reemplazarse por "¿Existen computadoras digitales imaginables que desempeñen bien el juego de la imitación?"*. No importa si las máquinas piensan o no, sino lo que importa es su habilidad para “engañar” a las personas, a través de la imitación. En 1960 el tema preferido era imitar un psicólogo, cuya interacción se reducía a elaborar preguntas. Hoy día tenemos “chatbots” o robots para conversación, en dominios muy específicos y contexto empresarial, que normalmente decepcionan a la gran mayoría de los clientes.

¿Ser exitosos en el juego de la simulación, permite concluir que la máquina “piensa”? Para los desarrolladores que los construyen, la máquina está simulando el comportamiento humano, pero es claro que no piensa, no es autónoma en su comportamiento, y no toma decisiones para las que no está programada a pesar de incluir algoritmos de aprendizaje automático. La computación avanzada que incluirá cada vez más elementos de interacción con el medio ambiente y con las personas, seguirá siendo el ejecutor de simulaciones previamente definidas. Los elementos de interacción permitirán a las máquinas (i) identificar su ubicación espacial (mejor que las personas), (ii) medir su nivel de carga de potencia, (iii) prever obstáculos que podrían ser objetos en movimiento y evitarlos,

teniendo autonomía física y de desplazamiento.

Hoy día las aspiradoras automáticas cumplen con estas tres características, al igual que los vehículos autónomos. Sin embargo, nadie califica la aspiradora o el vehículo como "ser pensante" y su autonomía no nos preocupa porque sabemos que es limitada. Con el avance exponencial de capacidades de cómputo, en un futuro cercano habrá más máquinas con estas características de interacción, y estaremos más acostumbrados a su usos y limitaciones.

4. "La IA fomenta una crisis sobre el ser humano y su esencia"
Para muchas personas enfrentar que hay máquinas más capaces que ellos o "más inteligentes" (por ganar una partida de ajedrez o por "deducir" o encontrar un diagnóstico médico) le puede generar crisis de diversos tipos. La primera crisis que afecta a la mayoría de las personas en la sociedad actual, es que su trabajo sea reemplazado por una máquina. El correo electrónico impactó a los carteros en bicicleta, así como los drones y los vehículos autónomos, podrían impactar los actuales mecanismos de distribución. Conocí muchos proyectos de automatización de información que justificaban su ROI (retorno de la inversión) con la reducción en los costos laborales, la cual no se daba debido a que las capacidades del nuevo personal a contratar eran a su vez más costosas, al pasar de auxiliares a analistas. Sin embargo, el resultado neto para la organización era un mayor rendimiento y mayores ingresos, normalmente. En el caso industrial, la automatización de plantas de producción, también generó mayores volúmenes y mejores ingresos, reduciendo el costo de los equipos producidos, que así pudieron llegar a más personas.

Igualmente hay autores que proyectan escenarios catastróficos en los cuales las máquinas "dominan" el mundo y "someten" a personas. El imaginario popular de una guerra atómica iniciada por líderes con "botones rojos" y de manera inadvertida, ha evolucionado a disparadores electrónicos, iniciados por máquinas y por motivos "de poder" asociados a intereses personales, o similares, que pudiesen tener las máquinas. El autor reitera que los computadores eminentemente **simulan** comportamientos humanos, para los que están previamente programados. Los robots construidos con características faciales humanas, están en capacidad de reír, llorar, expresar disgusto o sorpresa, de acuerdo a una programación predefinida incluyendo imitar las personas de ambientes compartidos actuales o anteriores (aprendizaje automático), pero eso no significa que la máquina tenga sentimientos.

5. "Tenemos que controlar la IA"
¿Puede "controlarse" el desarrollo de la tecnología? ¿Pueden "con-

trolarse" los resultados de estos desarrollos de la computación avanzada que hoy llamamos IA? La afirmación y estas preguntas se originan en calificar la tecnología de autónoma y eventualmente dañina para la sociedad.

Las herramientas son máquinas de apoyo a las personas, y no podemos calificar a estas herramientas de "malas" o contraproducentes. Son las personas que las utilizan mal, las que generan el daño. Una herramienta no es "buena" ni "mala" per se.

Solo por resaltar un ejemplo, Future for Life Institute (ver site, marzo 22 de 2023) es una organización con la misión de "dirigir la tecnología transformadora hacia el beneficio de la vida y alejarla de los riesgos extremos a gran escala", solicitó en marzo 2023 detener por seis meses los desarrollos en IA. Entre los firmantes de su misiva está incluido Elon Musk, fundador de OpenAI en 2015, y de xAI (startup de inteligencia artificial fundada el 12 de julio de 2023). Persiste la duda de los motivos y razones que pueden tener investigadores y fundadores de empresas de IA, pidiéndoles a sus colegas que ellos interrumpan sus trabajos y avances.

El autor llama la atención sobre los efectos de detener desarrollos en tecnología como la IA o la computación avanzada, hoy día presente en drones y equipos anti-drones, en satélites que evalúan y pronostican el curso de los huracanes, o que ayudan a evaluar la situación de cultivos de productos naturales en gran escala, o en equipos que son robots de ayuda en medicina, entre muchos otros ejemplos. Hay muchos sectores y personas beneficiados con la computación avanzada y detenerla porque existen riesgos, es perjudicial para la sociedad.

El futuro previsible de la computación avanzada y la IA

1. La metadata

El autor resalta la importancia de la metadata en todos los objetos generados, creados o manipulados con ayuda de herramientas de IA o computación avanzada. Debe ser obligatorio en el futuro cercano identificar si una foto o un mensaje fueron producidos o alterados por medios computacionales (IA) o corresponden a un autor humano identificable y localizable. Si recibimos una noticia, debe ser obligatorio poder identificar de manera específica su fuente, y la ubicación en la que se generó. Si un objeto computacional ha sido modificado o alterado varias veces, debe proporcionarse a todos los que lo consultan o reciben, la trazabilidad de la producción original y de las alteraciones. El uso de una tecnología como blockchain para almacenar esta trazabilidad, ayuda a certificar su fiabilidad.

La metadata de los objetos disminuirá la producción de noticias fal-

sas, ayudando a identificar las manipulaciones tendenciosas, y a sus autores.

Se recomienda que esta metadata siga estándares, previo un acuerdo de los protagonistas de la tecnología y que todos dispongamos de herramientas para su lectura y certificación.

2. Uso del contexto en el aprendizaje automático
El contexto de una conversación y el conocimiento de quién la origina son elementos claves y requeridos para lograr una traducción automática, o una comunicación productiva. Es un concepto básico que no parece ser considerado en muchas herramientas de comunicación como bots o en traducción automática, razón por la que este autor lo resalta.

3. Sobre la regulación al desarrollo y al uso de la IA
En la experiencia del autor la legislación actual es amplia y suficiente para identificar usos ilegales de las herramientas computacionales.

Atentar contra la privacidad de las personas, calumniar, suplantar, difundir noticias falsas, y muchos otros delitos no requieren de normas especiales para que estos sean identificados y que la justicia actúe. Se requiere con seguridad capacitar al personal de la justicia en cuanto a la caracterización de las herramientas computacionales que puedan ser utilizadas en la comisión de delitos, así como en definir el grado de responsabilidad de los que intervienen.

Una de las características de los avances de la computación ha sido su no restricción desde el punto de vista legal o normativo, lo que ha permitido un rápido desarrollo de la tecnología en general. Limitar su uso debido a los temores que genera el desconocimiento o al riesgo de que la tecnología sea mal utilizada, afectará el desarrollo y avance de la sociedad.

Reflexión final: La organización inteligente

El autor resalta que más importante que el desarrollo de la tecnología es su adaptación y correcta utilización, en nuestro entorno. Tener herramientas de IA, no garantiza que nuestras organizaciones y empresas no cometan errores computacionales ni errores en el manejo de la información. Podemos tener bots conversacionales, pero ¿tiene sentido (con o sin ellos) solicitar la identificación y autenticarla, varias veces en una interacción automatizada con un usuario? ¿Si un cliente tiene una sola cuenta o tarjeta, es inteligente que la aplicación con la que está interactuando pregunte qué cuenta desea manipular? ¿Cuántas veces damos repetidamente la nacionalidad o datos que son inmodificables, al generar un pasabordo con una aerolínea? ¿La organización que tiene nuestros documentos digitalizados, y probablemente son ellos mismos los

que los generan, nos solicita que los llevemos en papel?

Más importante que las herramientas es el uso que le damos en nuestro entorno y más en temas de nuestra responsabilidad.

Referencias

Buchanan, B.G. y Shortliffe, E.H. (1984). *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison-Wesley Editores.

Feigenbaum, Edward A. (1992). *Expert Systems: Principles and Practice*. The Encyclopedia of Computer Science and Engineering.

Harmon, Paul; King, David. (1985). *Expert Systems. Artificial Intelligence in Business*. Wiley Press.

Le Scao, Teven. (2020). *A brief history of machine translation paradigms*. https://medium.com/huggingface/a-brief-history-of-machine-translation-paradigms-d5c09d8a5b7e

Moor, James H. (Editor). (1989). *The Turing Test: The Elusive Standard of Artificial Intelligence (Studies in Cognitive Systems, 30).* Springer Ed.

Preis, Simon J. Ph.D. (Marzo 2023). *Are Expert Systems Dead? A review of recent trends, use cases and technologies*. https://towardsdatascience.com/are-expert-systems-dead

Thompson, Clive. (2021). *13 Milestones in the History of Robotics*. https://www.aventine.org/robotics/history-of-robotics

Turing, A. M. (1950). *Computing Machinery and Intelligence*. Mind Journal,

Future of Life Institute (Marzo 22 de 2023). *"Pause Giant AI Experiments: An Open Letter",* https://futureoflife.org/open-letter/pause-giant-ai-experiments/

Reason Why Madrid (2023). *Geoffrey Hinton, considerado ´padrino de la inteligencia artificial´, deja Google y advierte de los peligros de esta tecnología*. https://www.reasonwhy.es/actualidad/geoffrey-hinton-deja-google-advierte-peligros-inteligencia-artificial

IBM (s.f.) *"Deep Blue"*. Heritage. https://www.ibm.com/history/deep-blue

Julio E. López Medina, Ph.D., PMP. *Ingeniero de Sistemas, Doctorado en Informática del INPG de Grenoble, Francia, y su trabajo de tesis en el área de Inteligencia Artificial, se concentró en análisis semántico del lenguaje natural. Certificado como PMP® o "Project Management Profesional" por el PMI® Project Management Institute, Ha sido gerente de sistemas y de operaciones en el sector financiero y en multinacionales. Perito Técnico en Sistemas para Tribunales de Arbitramento de la Cámara de Comercio de Bogotá, principalmente.*

IA Generativa y Ciberseguridad

Gobernanza, riesgos emergentes y oportunidades para un futuro digital responsable

DOI: 10.29236/sistemas.n177a11

Resumen

La Inteligencia Artificial Generativa (GenAI) se ha consolidado como el catalizador más influyente de la transformación digital contemporánea. Su capacidad para producir contenido original, automatizar análisis complejos y expandir las capacidades cognitivas humanas está redefiniendo procesos empresariales, educativos y gubernamentales. Sin embargo, esta tecnología también reconfigura el panorama de amenazas: informes recientes, como el Microsoft Digital Defense Report 2025 (Microsoft, 2025), el IBM Threat Intelligence Index 2025 (IBM, 2025) y el ENISA Threat Landscape 2025 (ENISA, 2025), evidencian un repunte en ataques hiperpersonalizados, deepfakes operativos y variantes de malware asistidas por IA. Paradójicamente, la GenAI también es la clave para fortalecer los mecanismos de defensa, pues acelera la detección de anomalías y optimiza la respuesta ante incidentes. Este artículo analiza esa dualidad bajo la óptica del NIST Cybersecurity Framework 2.0 (NIST, 2024), con énfasis en la función Govern como base de la gobernanza algorítmica. Asimismo, se examina el uso ético de la tecnología en la academia y la empresa, junto con su impacto en la brecha digital global. Finalmente, se proponen líneas de investigación críticas para un entorno donde la velocidad del cambio exige decisiones informadas.

Palabras clave

IA Generativa, Ciberseguridad, Gobernanza, Responsabilidad, Resiliencia.

Introducción

La Inteligencia Artificial Generativa ha dejado de ser una novedad para convertirse en el núcleo del ecosistema digital actual. Los modelos fundacionales, capaces de generar texto, código, audio y video, han permeado todos los sectores y han alterado no solo la operación diaria, sino la esencia misma de los modelos de negocio. Accenture (Accenture, 2024) describe este fenómeno como la "reinvención del núcleo digital", un escenario donde la IA evoluciona de ser una herramienta de soporte a un amplificador sustancial de las capacidades humanas.

No obstante, este avance ocurre en un terreno hostil donde los riesgos mutan a la misma velocidad. Los actores maliciosos instrumentalizan la GenAI para automatizar la ingeniería social, generar contenidos sintéticos indistinguibles de la realidad y escalar sus operaciones con una eficiencia sin precedentes. Investigaciones recientes confirman una aceleración de ataques que explotan sesgos cognitivos humanos e introducen desinformación a escala industrial (Microsoft, 2025).

Simultáneamente, las organizaciones despliegan GenAI para enriquecer el análisis de seguridad, contextualizar amenazas y reducir los tiempos de respuesta. Este equilibrio dinámico entre riesgo y oportunidad demanda nuevas estrategias de gobernanza, ética y formación. El presente artículo aborda esta complejidad desde una perspectiva integral que entrelaza tecnología, educación y equidad digital.

Oportunidades y riesgos: La doble cara de la GenAI

La incorporación de la IA Generativa marca un punto de inflexión en la ciberseguridad al modificar drásticamente los tiempos, la escala y la sofisticación tanto del ataque como de la defensa.

La perspectiva ofensiva

La GenAI ha reducido significativamente las barreras de entrada para campañas avanzadas. Los adversarios ahora automatizan tareas críticas de la cadena de ataque:

- **Ingeniería social a escala:** Generación de *phishing* hiper-personalizado que imita la sintáxis y el tono de directivos o entidades de confianza.
- **Engaño sintético:** Uso de *deepfakes* de audio y video para manipular decisiones corporativas, eludir verificaciones biométricas o erosionar la reputación institucional (IBM, 2025).
- **Desarrollo de amenazas:** Aunque los modelos comerciales poseen salvaguardas, la IA ayuda indirectamente a explorar vulne-

rabilidades y crear variantes de malware polimórfico, lo que acelera el desarrollo de Tácticas, Técnicas y Procedimientos (TTPs).

La perspectiva defensiva

En contraparte, la defensa experimenta mejoras sustanciales en eficiencia analítica:

- **SOC Aumentados:** Los Centros de Operaciones de Seguridad procesan volúmenes masivos de datos para detectar patrones sutiles que escaparían al análisis humano tradicional.
- **Respuesta y recuperación:** La GenAI facilita la priorización de alertas, sugiere acciones de contención y sistematiza la documentación posterior al incidente, liberando al talento humano para tareas de mayor valor estratégico (Accenture, 2025).
- Esta naturaleza dual subraya la urgencia de fortalecer los controles y da paso al análisis de los marcos de referencia que deben guiar esta adopción.

Gobernanza algorítmica y el NIST CSF 2.0

El Cybersecurity Framework 2.0 del *National Institute of Standards and Technology* (NIST, 2024) posiciona a la función *Govern* (Gobernar) en el centro de su modelo y enfatiza que la ciberseguridad es una responsabilidad estratégica de liderazgo. En la era de la GenAI, esta función se vuelve vital para orquestar una arquitectura de decisiones que gestione el "riesgo algorítmico".

Bajo la función *Govern*, las organizaciones deben:

1. Instituir políticas de IA Responsable que definan el uso aceptable y ético.
2. Delimitar roles claros a lo largo del ciclo de vida de los modelos.
3. Evaluar vectores de ataque específicos, como los documentados en el marco ATLAS de MITRE (MITRE, 2023), que incluyen envenenamiento de datos, extracción de modelos e inyección de prompts.
4. Asegurar la supervisión humana (*Human-in-the-loop*) en decisiones críticas.
5. Asegurar la trazabilidad y auditoría de los resultados automatizados.

La GenAI potencia transversalmente las demás funciones del marco: mejora la clasificación de activos en *Identify*, endurece configuraciones en *Protect*, afina la sensibilidad en *Detect*, y agiliza la comunicación en *Respond* y *Recover*. Sin embargo, sin el eje rector de *Govern*, estas mejoras carecen de sostenibilidad y seguridad.

Ética y responsabilidad: Academia y Empresa

La IA Generativa ofrece un potencial inmenso, pero introduce desafíos éticos y cognitivos que no pueden ignorarse.

En el **ámbito académico**, la GenAI puede democratizar el acceso a tutorías personalizadas y explicar conceptos complejos. Sin embargo, organismos como la UNESCO (UNESCO, 2023) y la OCDE (OECD, 2023) advierten sobre el riesgo de atrofia en el pensamiento crítico y la dificultad para verificar la autoría. Un uso responsable exige rediseñar las evaluaciones para privilegiar el razonamiento humano, la creatividad y la argumentación, habilidades que la IA aún no puede replicar con autenticidad.

En el **entorno empresarial**, la integración de GenAI impulsa la reinvención operativa. No obstante, esto requiere una gestión rigurosa para evitar sesgos algorítmicos, proteger la propiedad intelectual y evitar la fuga de datos en modelos públicos. La transparencia y la interpretabilidad (*Explainable AI*) son requisitos no negociables para mantener la confianza de clientes y empleados.

GenAI y la brecha digital global

La adopción de GenAI no ocurre en el vacío, sino en un contexto de desigualdad estructural. El Global Cybersecurity Outlook 2025 (WEF, 2025) destaca una bifurcación clara:

- **Países Desarrollados:** Cuentan con ecosistemas de innovación robustos, talento especializado y marcos regulatorios maduros, condiciones que les permiten liderar iniciativas de "IA Soberana".
- **Países Emergentes:** Enfrentan brechas de conectividad y una escasez crítica de talento en ciberseguridad, documentada por ISC2 (ISC2, 2025). La dependencia de proveedores tecnológicos externos aumenta su vulnerabilidad y limita su autonomía digital.

Pese a esto, la GenAI podría ser una herramienta para cerrar brechas si se utiliza estratégicamente en salud, agricultura y educación. El éxito dependerá de políticas públicas inteligentes, cooperación internacional y una inversión sostenida en infraestructura digital propia.

Conclusiones

La IA Generativa no es una tecnología más; se trata de un cambio de paradigma que redefine la seguridad, la educación y la gobernanza. Su impacto final dependerá de nuestra capacidad para adoptarla bajo principios de responsabilidad y estrategia. La función Govern del NIST (NIST, 2024) emerge como el pilar fundamental para gestionar esta transición.

La desigualdad digital plantea el reto más grande: asegurar que la GenAI funcione como un puente hacia el desarrollo y no como un muro que aísle a las economías emergentes. Frente a un horizonte donde lo sintético amenaza la confianza, nuestra respuesta defensiva debe ser **REAL**: basada en la **R**esponsabilidad de la gobernanza,

la **E**valuación continua de riesgos, una **A**rquitectura de seguridad robusta y la **L**ucidez humana para discernir la verdad. Solo combinando estos elementos transformaremos el riesgo algorítmico en una ventaja estratégica.

Interrogantes para el próximo año

- ¿Cómo evolucionarán los ataques impulsados por agentes totalmente autónomos (Accenture, 2025)?
- ¿Qué mecanismos de supervisión humana serán indispensables para preservar la confianza institucional?
- ¿Podrán los países emergentes desarrollar estrategias que reduzcan su dependencia tecnológica?
- ¿Qué estándares globales prevalecerán para garantizar un uso transparente de la GenAI?

Líneas de investigación futura sugeridas

- Modelos de gobernanza algorítmica adaptados a países de baja madurez institucional.
- Implementación de prácticas MLSecOps para asegurar el ciclo de vida completo de la IA.
- Desarrollo de herramientas de auditoría algorítmica asistidas por la propia GenAI.
- Impacto de los agentes autónomos en infraestructuras críticas.
- Evolución del talento en ciberseguridad frente a la automatización.

Referencias

Accenture. (2024, January 16). *Securing the digital core: Elevating cybersecurity for the AI-driven enterprise*. Retrieved from Accenture: https://www.accenture.com/us-en/insights/cybersecurity/securing-digital-core

Accenture. (2025, July 10). *Empowering a secure autonomous AI future*. Retrieved from Accenture Security Blog: https://www.accenture.com/us-en/blogs/security/empowering-secure-autonomous-ai-future

Accenture. (2025, June 26). *State of Cybersecurity Resilience 2025*. Retrieved from Accenture: https://www.accenture.com/us-en/insights/security/state-cybersecurity-2025

ENISA. (2025, October 01). *ENISA Threat Landscape 2025*. Retrieved from ENISA: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2025

IBM. (2025, February 24). *X-Force Threat Intelligence Index 2025*. Retrieved from IBM: https://www.ibm.com/reports/threat-intelligence

ISC2. (2025, September 09). *2025 Cybersecurity Hiring Trends: Skills Deep Dive*. Retrieved from ISC2: https://www.isc2.org/Insights/2025/09/cybersecurity-hiring-trends-skills-deep-dive

Microsoft. (2025, October 16). *Microsoft Digital Defense Report 2025*. Retrieved from Microsoft Security Insider: https://www.microsoft.com/en-us/security/security-insider/intelligence-reports/microsoft-digital-defense-report-2025

MITRE. (2023, August 01). *ATLAS: Adversarial Threat Landscape for AI Systems*. Retrieved from MITRE ATLAS: https://atlas.mitre.org/

NIST. (2024, February 26). *The NIST Cybersecurity Framework (CSF) 2.0*. Retrieved from NIST: https://www.nist.gov/cyberframework

OECD. (2023, November 22). *OECD Digital Education Outlook 2023: Global Standards for AI in Education*. Retrieved from OECD iLibrary: https://www.oecd.org/en/publications/oecd-digital-education-outlook-2023_c74f03de-en.html

UNESCO. (2023, September 07). *Guidance for generative AI in education and research*. Retrieved from UNESCO Digital Library: https://unesdoc.unesco.org/ark:/48223/pf0000386693

WEF. (2025, January 13). *Global Cybersecurity Outlook 2025.* Retrieved from World Economic Forum: https://www.weforum.org/publications/global-cybersecurity-outlook-2025/

Juan Mario Posada *es Senior Manager en Accenture con más de 20 años de experiencia en ciberseguridad IT/OT y gestión de riesgos en Latinoamérica y EE. UU. Especialista en sectores críticos, cuenta con certificaciones como CISSP, CISM, CDPSE, CRISC y CISA. Ha asesorado empresas a nivel regional y global en la transformación estratégica de programas de ciberseguridad. Actualmente lidera la innovación y expansión de servicios de seguridad ciberfísica para Latinoamérica desde Houston.*

ACIS

ASOCIACIÓN COLOMBIANA DE INFORMÁTICA, SISTEMAS Y TECNOLOGÍAS AFINES

¡AFILIATE YA!

Y DISFRUTA DE ESTOS BENEFICIOS

- Actualización en formación profesional y académica de manera constante.
- Candidatura a participación profesional en proyectos de ACIS.
- Candidato a Director o CoDirector de Grupo de Interés (GI).
- Candidatura a participar en consultorías solicitadas a ACIS por el sector privado y público.
- Candidatura a participar en eventos nacionales o internacionales como delegado de ACIS.
- Candidato a Miembro de Consejo Editorial de la Revista Sistemas.
- Descuentos especiales en cursos y eventos exclusivos en el área de las TIC.
- Referencia profesional para vinculación como Perito en procesos de arbitraje.
- Referencia para participación en Juntas Directivas.
- Inclusión en el gremio de Ingenieros de Sistemas más importante del país.
- Recepción trimestral de la revista SISTEMAS en formato digital.
- Acceso diferido a la base de Webinars de ACIS.
- Acceso exclusivo a oportunidades laborales a través de nuestro portal de empleo.
- Participación como conferencista o participante en las charlas semanales.
- Correo personal con @acis.org.co
- Asista a las funciones del Teatro Nacional con un 20% de descuento. Consulte la Programación y solicite el descuento a cursos@acis.org.co.
- 30% de descuento en los libros de la Casa Editorial ALFAOMEGA, consulte el Catálogo

Afiliación General

Afiliación + Precio de estudio de formulario

$ 323.400 + $ 60.000

Afiliación para recién egresados

Descuento de 20% para recién egresados, (2 años) + Precio de estudio de formulario

$ 258.800 + $ 60.000

301 5530540 | suscripciones@acis.org.co | www.acis.org.co

ACIS

ASOCIACIÓN COLOMBIANA DE INFORMÁTICA, SISTEMAS Y TECNOLOGÍAS AFINES

¡ Pongase al día en sus cuotas !

Recuerda que te da derecho a participar en un evento virtual

Comuniquese con nuestro equipo de atención al cliente.
suscripciones@acis.org.co
o al télefono 3015530540

Para más información
www.acis.org.co/

CONECTA CON NOSOTROS

@Comunidadacis

ACIS
ASOCIACIÓN COLOMBIANA DE INFORMÁTICA, SISTEMAS Y TECNOLOGÍAS AFINES

www.acis.org.co