RCMOS – Revista Científica Multidisciplinar O Saber.
ISSN: 2675-9128. São Paulo - SP.
Año VII, vol. 1, 2026 | Envío: 3 de julio de 2026 | Aceptación: 6 de julio de 2026 | Publicación: 9 de julio de 2026

Derecho y tecnología en diálogo: un análisis conjunto de los impactos de la inteligencia artificial generativa en la privacidad y la seguridad de la información en el contexto corporativo

Direito e tecnologia em diálogo: uma análise conjunta dos impactos da inteligência artificial generativa na privacidade e segurança da informação no contexto corporativo
Law and Technology in Dialogue: A Joint Analysis of the Impacts of Generative Artificial Intelligence on Privacy and Information Security in the Corporate Context

Manoel Pereira de Lima Junior
Juliana Beatriz Moraes Miura

Resumen:
La rápida expansión de la Inteligencia Artificial Generativa (IAG) en el entorno corporativo impone nuevos desafíos a la protección de datos personales, la seguridad de la información y la gobernanza organizacional. El presente artículo analiza, desde una perspectiva interdisciplinaria, los impactos jurídicos, tecnológicos y filosóficos derivados del uso de plataformas de IAG por las organizaciones, con énfasis en la Ley General de Protección de Datos de Brasil (LGPD – Ley n.º 13.709/2018), en los riesgos de filtración de información confidencial y en los desafíos del fenómeno denominado Shadow AI. La investigación articula fundamentos del Derecho Digital con principios de Seguridad de la Información, normas técnicas reconocidas internacionalmente (ISO/IEC 27001, ISO/IEC 42001, NIST AI RMF, OWASP LLM Top 10) y referentes filosóficos de Michel Foucault, Zygmunt Bauman, Byung-Chul Han y Shoshana Zuboff. Como resultado, se demuestra que la protección efectiva de los activos informacionales corporativos exige un enfoque integrado entre personas, procesos, tecnología y cultura organizacional, superando la dicotomía entre innovación tecnológica y conformidad jurídica. Se propone un modelo de gobernanza responsable de la IAG fundamentado en principios de Privacy by Design, Zero Trust, gestión de riesgos y ética digital.

Palabras clave:
Inteligencia Artificial Generativa. LGPD. Seguridad de la Información. Shadow AI. Gobernanza de Datos. Privacidad. Ciberseguridad Corporativa.

Resumo:
A rápida expansão da Inteligência Artificial Generativa (IAG) no ambiente corporativo impõe novos desafios à proteção de dados pessoais, à segurança da informação e à governança organizacional. O presente artigo analisa, sob perspectiva interdisciplinar, os impactos jurídicos, tecnológicos e filosóficos decorrentes do uso de plataformas de IAG pelas organizações, com ênfase na Lei Geral de Proteção de Dados (LGPD – Lei nº 13.709/2018), nos riscos de vazamento de informações confidenciais e nos desafios do fenômeno denominado Shadow AI. A investigação articula fundamentos do Direito Digital com princípios de Segurança da Informação, normas técnicas reconhecidas internacionalmente (ISO/IEC 27001, ISO/IEC 42001, NIST AI RMF, OWASP LLM Top 10) e referenciais filosóficos de Michel Foucault, Zygmunt Bauman, Byung-Chul Han e Shoshana Zuboff. Como resultado, demonstra-se que a proteção efetiva dos ativos informacionais corporativos exige uma abordagem integrada entre pessoas, processos, tecnologia e cultura organizacional, superando a dicotomia entre inovação tecnológica e conformidade jurídica. Propõe-se um modelo de governança responsável da IAG fundamentado em princípios de Privacy by Design, Zero Trust, gestão de riscos e ética digital.

Año VII, vol. 1, 2026 | Envío: 3 de julio de 2026 | Aceptación: 6 de julio de 2026 | Publicación: 9 de julio de 2026

Palavras-chave:
Inteligência Artificial Generativa. LGPD. Segurança da Informação. Shadow AI. Governança de Dados. Privacidade. Cibersegurança Corporativa.

Abstract:
The rapid expansion of Generative Artificial Intelligence (GAI) in the corporate environment poses new challenges to personal data protection, information security, and organizational governance. This article analyzes, from an interdisciplinary perspective, the legal, technological, and philosophical impacts arising from the use of GAI platforms by organizations, with emphasis on Brazil's General Data Protection Law (LGPD – Law No. 13,709/2018), the risks of confidential information leakage, and the challenges posed by the phenomenon known as Shadow AI. The investigation articulates principles of Digital Law with Information Security standards, internationally recognized technical norms (ISO/IEC 27001, ISO/IEC 42001, NIST AI RMF, OWASP LLM Top 10), and philosophical frameworks drawn from Michel Foucault, Zygmunt Bauman, Byung-Chul Han, and Shoshana Zuboff. As a result, it is demonstrated that the effective protection of corporate informational assets requires an integrated approach encompassing people, processes, technology, and organizational culture, transcending the dichotomy between technological innovation and legal compliance. A model of responsible GAI governance is proposed, grounded in Privacy by Design, Zero Trust, risk management, and digital ethics.

Keywords:
Generative Artificial Intelligence. LGPD. Information Security. Shadow AI. Data Governance. Privacy. Corporate Cybersecurity.

INTRODUCCIÓN

La Inteligencia Artificial Generativa (IAG) [1]se ha consolidado como una de las innovaciones tecnológicas más disruptivas del siglo XXI. Los Modelos de Lenguaje a Gran Escala (MLG) , como GPT-4 de OpenAI, Gemini de Google y Claude de Anthropic , se han vuelto accesibles al público en general y han penetrado rápidamente en los entornos corporativos, transformando los procesos de trabajo, la comunicación organizacional, el análisis legal, el desarrollo de software y la gestión estratégica (BOMMASANI et al., 2021).

Esta penetración tecnológica, si bien aporta beneficios innegables en productividad e innovación, conlleva riesgos que trascienden las dimensiones puramente técnicas. En el ámbito jurídico, el uso de plataformas públicas de información, educación y comunicación (IEC) por parte de empleados y organizaciones plantea interrogantes fundamentales sobre el tratamiento de datos personales, la responsabilidad civil y administrativa de los responsables del tratamiento

[1]El término «Inteligencia Artificial Generativa» (IAG) se refiere a los sistemas de IA capaces de producir contenido nuevo (texto, imágenes, audio, código) a partir de patrones aprendidos en grandes volúmenes de datos. Se distingue de la IA discriminativa, que clasifica o predice categorías en datos existentes.

Año VII, vol. 1, 2026 | Envío: 3 de julio de 2026 | Aceptación: 6 de julio de 2026 | Publicación: 9 de julio de 2026

y el cumplimiento del marco normativo brasileño para la protección de datos: la Ley General de Protección de Datos (LGPD), Ley n.º 13.709, del 14 de agosto de 2018 (BRASIL, 2018). Este marco normativo se inspiró en el Reglamento General de Protección de Datos (RGPD) europeo [2]y constituye el principal instrumento jurídico para la protección de la privacidad en Brasil.

En el ámbito de la seguridad de la información, la adopción no controlada de herramientas de IA por parte de los empleados —un fenómeno conocido como IA en la sombra— representa una amenaza para la confidencialidad, la integridad y la disponibilidad de los activos de información de las organizaciones. Datos estratégicos, secretos comerciales, información de clientes y propiedad intelectual pueden insertarse inadvertidamente en modelos entrenados y mantenidos por terceros, con consecuencias potencialmente irreversibles (GARTNER, 2023).

Ante este panorama, este artículo plantea el siguiente problema de investigación: ¿cómo pueden las organizaciones equilibrar los beneficios que ofrece la Inteligencia Artificial Generativa con la necesidad de protección legal de datos, seguridad de la información y preservación de sus activos estratégicos?

La justificación de esta investigación radica en la urgencia y relevancia del tema. Brasil, a pesar de contar con una legislación de protección de datos comparable en alcance al RGPD europeo, aún carece de regulaciones específicas sobre el uso de la inteligencia artificial, lo que hace indispensable un análisis integral de los marcos regulatorios existentes y las mejores prácticas internacionales. La ejecución de estas tareas recae principalmente en la Autoridad Nacional de Protección de Datos (ANPD) [3], que ha ido ampliando gradualmente su función regulatoria en este ámbito.

El objetivo general de esta investigación es analizar los impactos legales, tecnológicos y filosóficos de la Inteligencia Artificial Generativa en la privacidad y la seguridad de la información en el contexto corporativo. Los objetivos específicos son: (a) examinar el marco regulatorio de la LGPD (Ley General de Protección de Datos de Brasil) y su aplicación al uso de la IA; (b) identificar los principales riesgos de seguridad de la información asociados con el

[2]El RGPD (Reglamento General de Protección de Datos) El Reglamento (UE) 2016/679 entró en vigor el 25 de mayo de 2018 y se considera la referencia mundial en materia de protección de datos personales, habiendo influido directamente en la elaboración de la LGPD brasileña.

[3]La ANPD (Autoridad Nacional de Protección de Datos) fue creada mediante el Decreto n.º 10.474/2020 como un organismo autónomo con fines especiales, vinculado a la Presidencia de la República, dotado de autonomía técnica y decisoria, patrimonio propio e ingresos específicos. Es responsable de garantizar la protección de datos personales, elaborar directrices para la Política Nacional de Protección de Datos y supervisar y aplicar sanciones.

Año VII, vol. 1, 2026 | Envío: 3 de julio de 2026 | Aceptación: 6 de julio de 2026 | Publicación: 9 de julio de 2026

uso corporativo de la IA; (c) analizar el fenómeno de la IA en la sombra y sus desafíos de gobernanza; (d) articular los fundamentos filosóficos sobre poder, datos y vigilancia en el contexto de la IA; y (e) proponer directrices para una gobernanza responsable de la IA en el entorno corporativo.

La metodología adoptada es bibliográfica, exploratoria y cualitativa, basada en una revisión de la literatura científica, la legislación aplicable, las normas técnicas y los documentos institucionales, tal como se describe en la sección metodológica específica.

Evolución de la inteligencia artificial generativa en el entorno empresarial.

La trayectoria de la Inteligencia Artificial abarca décadas de avances científicos y cambios de paradigma. Desde los primeros modelos simbólicos y sistemas expertos de las décadas de 1950 y 1960, pasando por las redes neuronales artificiales y el aprendizaje automático supervisado, llegamos al paradigma del aprendizaje profundo , que revolucionó las capacidades computacionales del procesamiento del lenguaje natural, la visión por computadora y la generación de contenido (GOODFELLOW; BENGIO; COURVILLE, 2016). El hito transformador para la IA fue la introducción de la arquitectura Transformer [4], propuesta por Vaswani et al. (2017) en el artículo fundamental " Atención ". Es Todo Tú Necesidad ".

Esta arquitectura permitió el entrenamiento de modelos a una escala sin precedentes, culminando en el desarrollo de LLM que caracterizan la generación actual de herramientas de IA. El proceso de ajuste fino a través del aprendizaje por refuerzo desde La retroalimentación humana (RLHF) [5]ha logrado que estos modelos se ajusten mejor a las expectativas de los usuarios. El lanzamiento público de ChatGPT en noviembre de 2022 representó un punto de inflexión: en menos de cinco días, la plataforma alcanzó el millón de usuarios, y en dos meses, los cien millones, una velocidad de adopción sin precedentes en la historia de la tecnología de consumo (OPENAI, 2023).

---

[4]modelo de arquitectura Transformer , introducido por Vaswani et al. (2017), utiliza mecanismos de atención (atención). Mecanismos que permiten al modelo ponderar la relevancia de diferentes partes del texto de entrada al generar cada token de salida. Esta arquitectura supera las limitaciones de las redes neuronales recurrentes (RNN) y permite una paralelización masiva del entrenamiento.

[5]El RLHF (Aprendizaje por refuerzo a partir de La retroalimentación humana (RLHF) es una técnica de ajuste fino para modelos de aprendizaje por refuerzo (LLM) en la que evaluadores humanos califican las respuestas del modelo. Esta calificación se utiliza para entrenar un modelo de recompensa que guía el aprendizaje por refuerzo del LLM. La RLHF ha sido fundamental para que modelos como ChatGPT se ajusten mejor a las expectativas y preferencias de los usuarios.

Año VII, vol. 1, 2026 | Envío: 3 de julio de 2026 | Aceptación: 6 de julio de 2026 | Publicación: 9 de julio de 2026

En el ámbito empresarial, la penetración de IAG se produjo en múltiples dimensiones. Los departamentos jurídicos comenzaron a utilizar herramientas de gestión de contratos para la revisión de contratos y la investigación de jurisprudencia; los equipos de tecnología adoptaron herramientas de generación de código; los departamentos de marketing exploraron la creación automatizada de contenido; y los gerentes utilizaron las herramientas para la síntesis de informes y el apoyo a la toma de decisiones (MCKINSEY GLOBAL INSTITUTE, 2023).

Según un informe del McKinsey Global Institute (2023), la IA tiene el potencial de automatizar actividades que representan entre el 60 % y el 70 % del tiempo laboral actual de los empleados en diversas funciones cognitivas. Sin embargo, este potencial transformador conlleva riesgos que la mayoría de las organizaciones aún no están preparadas para gestionar adecuadamente.

Gartner (2023) identificó que, para 2025, más del 80 % de las empresas habrán implementado aplicaciones de IA en sus procesos, pero menos del 30 % habrá establecido políticas y controles adecuados para mitigar los riesgos asociados. Esta asimetría entre la adopción y la gobernanza constituye el núcleo del problema que se analiza en este artículo.

Los datos como nuevo activo estratégico para las organizaciones

El auge de la economía digital ha transformado radicalmente la naturaleza y el valor de los datos. Anteriormente considerados un subproducto de las operaciones organizativas, los datos se han convertido en el principal activo estratégico de las empresas del siglo XXI. La expresión popularmente atribuida al economista británico Clive Humby —"los datos son ..." "El nuevo petróleo " resume, aunque de forma simplificada, la importancia que la información estructurada y no estructurada ha adquirido en la creación de valor económico (THE ECONOMIST, 2017).

En el ámbito empresarial, los datos se clasifican en múltiples categorías estratégicas: datos de clientes (historial de compras, preferencias, datos financieros), datos de empleados (información personal, datos de rendimiento, historial médico), propiedad intelectual (código fuente, fórmulas, procesos de producción), información financiera confidencial y secretos comerciales. Cada una de estas categorías representa, a la vez, una ventaja competitiva y una responsabilidad legal; su divulgación no autorizada puede generar importantes daños financieros, reputacionales y legales.

Valorar los datos como un activo estratégico tiene implicaciones directas para el gobierno corporativo. Los consejos de administración y la alta dirección han llegado a

Año VII, vol. 1, 2026 | Envío: 3 de julio de 2026 | Aceptación: 6 de julio de 2026 | Publicación: 9 de julio de 2026

reconocer que la protección de datos no es simplemente una cuestión técnica o de cumplimiento normativo, sino una dimensión fundamental de la estrategia empresarial y la gestión de riesgos (WORLD ECONOMIC FORUM, 2020).

Desde esta perspectiva, la llegada de IAG al entorno corporativo crea una paradoja: las mismas herramientas que permiten extraer valor de los datos con una eficiencia sin precedentes son, simultáneamente, vectores potenciales para la filtración y la exposición no autorizada de esta información. Cuando un empleado introduce datos corporativos confidenciales en una plataforma pública de IAG para obtener análisis, sugerencias o resúmenes, puede estar transfiriendo este activo estratégico a servidores de terceros, con consecuencias legales y competitivas potencialmente devastadoras.

Por lo tanto, la clasificación y el inventario de datos corporativos se convierten en requisitos indispensables para cualquier estrategia de gobernanza de la IA. Sin saber qué datos existen, dónde se encuentran y su nivel de confidencialidad, las organizaciones no pueden establecer controles adecuados sobre qué datos pueden o no ser procesados por herramientas externas de inteligencia artificial (ISO/IEC 27001, 2022).

ANÁLISIS JURÍDICO: LGPD, PRIVACIDAD Y RESPONSABILIDAD CORPORATIVA

La Ley General de Protección de Datos (LGPD), promulgada el 14 de agosto de 2018 y en vigor desde septiembre de 2020, constituye el principal marco normativo brasileño para el tratamiento de datos personales. Inspirada en el Reglamento General de Protección de Datos (RGPD) de la Unión Europea (Reglamento 2016/679), la LGPD establece principios, derechos de los titulares de los datos, obligaciones de los encargados del tratamiento y sanciones administrativas aplicables por el organismo regulador: la ANPD.

El artículo 6 de la LGPD (Ley General de Protección de Datos de Brasil) enumera los principios que deben guiar el tratamiento de datos personales, con especial relevancia para el contexto de la IAG: finalidad (tratamiento para fines legítimos, específicos y explícitos informados al interesado); adecuación (compatibilidad con las finalidades informadas); necesidad (limitación al mínimo necesario para la finalidad); libre acceso; calidad de los datos; transparencia; seguridad; prevención; no discriminación; y responsabilidad (BRASIL, 2018, art. 6).

En cuanto a las bases legales para el tratamiento de datos personales, el artículo 7 de la LGPD (Ley General de Protección de Datos de Brasil) enumera las circunstancias que autorizan

Año VII, vol. 1, 2026 | Envío: 3 de julio de 2026 | Aceptación: 6 de julio de 2026 | Publicación: 9 de julio de 2026

dicho tratamiento, entre las que destacan el consentimiento del interesado, el interés legítimo del responsable del tratamiento y el cumplimiento de una obligación legal. Para los datos personales sensibles [6]—categoría que incluye datos sobre salud, orientación sexual, creencias religiosas, datos genéticos y biométricos—, el artículo 11 de la LGPD impone restricciones aún más estrictas, exigiendo generalmente el consentimiento específico y explícito del interesado.

La inserción de datos personales en plataformas públicas de IAG sin una base legal adecuada, sin cláusulas contractuales específicas con los proveedores y sin una evaluación de impacto en la protección de datos puede constituir una infracción de la LGPD (Ley General de Protección de Datos de Brasil) sujeta a sanciones que incluyen advertencias, una multa simple de hasta el 2% de los ingresos de la empresa en Brasil (limitada a R$ 50 millones por infracción), multas diarias, divulgación pública de la infracción, bloqueo o eliminación de los datos personales involucrados (BRASIL, 2018, art. 52).

La responsabilidad civil de las organizaciones en el contexto de la Gestión Inteligente de Datos (GID) presenta dimensiones particularmente complejas. El artículo 42 de la LGPD (Ley General de Protección de Datos de Brasil) estipula que el responsable del tratamiento u operador que, debido al ejercicio de actividades de tratamiento de datos personales, cause daños patrimoniales, morales, individuales o colectivos al interesado, está obligado a repararlos. La cuestión de la distribución de la responsabilidad entre el responsable del tratamiento (la organización que define los fines del tratamiento) y el operador (el proveedor de la plataforma GID) es un tema que requiere especial atención en los contratos de servicios.

Desde esta perspectiva, el fenómeno de la IA en la sombra puede llevar a una organización a asumir el rol de controladora del procesamiento de datos sin siquiera ser consciente de ello. Esto subraya la necesidad de contar con políticas claras, mecanismos de control y programas internos de concienciación.

La transferencia internacional de datos representa otra dimensión legal crucial. Los principales proveedores de plataformas públicas de IA general (OpenAI, Google, Microsoft, Anthropic) son empresas estadounidenses, cuyos servidores se encuentran mayoritariamente fuera del territorio brasileño. El artículo 33 de la LGPD (Ley General de Protección de Datos de Brasil) condiciona la transferencia internacional de datos personales a la existencia de un nivel adecuado de protección en el país de destino (que debe ser reconocido por la ANPD -

[6]El artículo 5, apartado II, de la LGPD (Ley General de Protección de Datos de Brasil) define los datos personales sensibles como aquellos datos "relativos al origen racial o étnico, las creencias religiosas, las opiniones políticas, la afiliación a un sindicato u organización religiosa, filosófica o política, los datos relativos a la salud o la vida sexual, los datos genéticos o biométricos, cuando estén vinculados a una persona física".

Autoridad Nacional de Protección de Datos), la existencia de garantías contractuales específicas u otras hipótesis exhaustivas (ANPD, 2023).

La ANPD (Autoridad Nacional de Protección de Datos) ha avanzado en la publicación de directrices relevantes para el contexto de la IA. Sin embargo, la normativa específica sobre IA aún se encuentra en desarrollo dentro del proceso legislativo brasileño; en particular, el Proyecto de Ley N° 2338/2023 [7], que se está tramitando en el Senado Federal y propone un marco regulatorio específico para la inteligencia artificial en Brasil.

En el ámbito de la responsabilidad administrativa, la ANPD (Autoridad Nacional de Protección de Datos) puede iniciar procedimientos de sanción administrativa contra las organizaciones que utilicen plataformas de IA sin cumplir con los requisitos de la LGPD (Ley General de Protección de Datos de Brasil). Demostrar la adopción de medidas de seguridad técnicas y organizativas, así como la existencia de un programa sólido de gobernanza de la privacidad, son factores que pueden mitigar la responsabilidad de los responsables del tratamiento de datos (BRASIL, 2018, art. 50).

ANÁLISIS TECNOLÓGICO: RIESGOS, VULNERABILIDADES Y CONTROLES DE SEGURIDAD

Desde la perspectiva de la seguridad de la información, la inteligencia artificial generativa representa un vector de riesgo multidimensional que desafía los modelos tradicionales de protección corporativa. Para comprender adecuadamente estos riesgos, es necesario analizar el funcionamiento técnico de las herramientas de gestión del aprendizaje (LLM) y sus vulnerabilidades inherentes, así como los riesgos asociados al uso corporativo no controlado de estas herramientas.

(LLM) se entrenan con vastos conjuntos de datos textuales mediante técnicas de aprendizaje profundo basadas en la arquitectura Transformer . Una vez entrenados, estos modelos se ajustan con datos y técnicas adicionales, como RLHF, para adaptar su comportamiento a las expectativas del usuario. El proceso de inferencia —la generación de respuestas en tiempo real— se lleva a cabo en los servidores del proveedor, lo que significa que

[7]El proyecto de ley n.º 2338/2023, del senador Rodrigo Pacheco, propone un marco regulatorio integral para la inteligencia artificial en Brasil, basado en un enfoque de riesgo similar al de la Ley de IA de la UE . El proyecto de ley clasifica los sistemas de IA en tres categorías de riesgo —inaceptable, alto y limitado— y establece obligaciones diferenciadas para cada categoría, incluyendo evaluaciones de cumplimiento, transparencia y supervisión humana.

Año VII, vol. 1, 2026 | Envío: 3 de julio de 2026 | Aceptación: 6 de julio de 2026 | Publicación: 9 de julio de 2026

cada solicitud enviada por el usuario se transmite y procesa en la infraestructura de terceros (OUYANG et al., 2022).

En 2023, OWASP (Open Web Application Security Project) publicó el "OWASP Top 10 para aplicaciones de modelos de lenguaje a gran escala", un documento que cataloga las principales vulnerabilidades de seguridad de los sistemas basados en LLM . Las más relevantes para el contexto corporativo incluyen: inyección de indicaciones (manipulación del comportamiento del modelo mediante instrucciones maliciosas insertadas en las indicaciones); fuga de datos confidenciales contenidos en el contexto de la conversación; entrenamiento con datos envenenados (envenenamiento de datos); generación de contenido poco fiable —las llamadas "alucinaciones" [8]— con potencial para causar daño; y dependencia excesiva de los resultados del modelo sin una validación humana adecuada (OWASP, 2023).

En cuanto al riesgo de exposición de información confidencial, estudios realizados por empresas de seguridad como Cyberhaven y LayerX Security (2023) revelaron que una parte significativa de los datos introducidos por los empleados en plataformas públicas de IAG contiene información corporativa sensible. El episodio de Samsung [9]en marzo de 2023 se convirtió en un caso emblemático: ingenieros de la compañía insertaron código fuente propietario en ChatGPT para depurarlo, lo que provocó la posible exposición de secretos comerciales (BUSINESSINSIDER, 2023).

La norma ISO/IEC 27001:2022, que especifica los requisitos para establecer, implementar, mantener y mejorar continuamente un Sistema de Gestión de la Seguridad de la Información (SGSI), proporciona el marco estructural para gestionar estos riesgos. El Anexo A de la norma enumera los controles de seguridad organizados en cuatro categorías: organizativos, humanos, físicos y tecnológicos, que, aplicados al contexto de la IAG, implican la necesidad de políticas de uso aceptable, capacitación de los empleados, evaluación de proveedores y controles técnicos para la prevención de la pérdida de datos. Prevención (DLP) y gestión de vulnerabilidades específicas en aplicaciones basadas en IA (ISO/IEC 27001:2022).

---

[8]El fenómeno de las "alucinaciones" en los modelos de lógica descriptiva (MLD) se refiere a la generación de información fáctica incorrecta, referencias bibliográficas inexistentes o afirmaciones no verificables presentadas con aparente seguridad por el modelo. Este fenómeno es consecuencia de la naturaleza estadística de los MLD , que generan tokens basados en probabilidades, sin acceso a un modelo de verdad interno ni a la verificación fáctica.

[9]En marzo de 2023, empleados de Samsung Semiconductor insertaron código fuente propietario en ChatGPT en al menos tres ocasiones distintas: para depurar código de base de datos, para verificar código relacionado con equipos de semiconductores y para convertir notas de reuniones en presentaciones. Posteriormente, Samsung prohibió el uso de herramientas externas de IA generativa en sus dispositivos corporativos.

RCMOS – Revista Científica Multidisciplinar O Saber.
ISSN: 2675-9128. São Paulo - SP.
Año VII, vol. 1, 2026 | Envío: 3 de julio de 2026 | Aceptación: 6 de julio de 2026 | Publicación: 9 de julio de 2026

La norma ISO/IEC 42001:2023 [10]complementa este marco con requisitos específicos para sistemas de gestión de inteligencia artificial.

La arquitectura de confianza cero [11], formalizada por el NIST en la publicación especial SP 800-207 (2020), se basa en el principio de que ningún usuario, dispositivo o servicio debe considerarse de confianza por defecto. Aplicado al contexto de la gestión de la identidad y la comunicación (GIC), esto implica que el acceso a las herramientas de GIC debe otorgarse en función de políticas de autorización explícitas, verificación continua de la identidad y monitorización del comportamiento, con los privilegios mínimos necesarios para la función que se realiza.

Soluciones de agente de seguridad de acceso a la nube (CASB) [12]y de prevención de pérdida de datos La prevención de pérdida de datos (DLP) completa el arsenal técnico disponible para mitigar los riesgos de la IA empresarial. Finalmente, el despliegue de entornos privados y seguros para la IA empresarial —como instancias locales o en la nube privada de modelos de lenguaje— permite a las organizaciones disfrutar de los beneficios de la IA sin exponer datos confidenciales a terceros (MICROSOFT, 2023).

La IA en la sombra y los nuevos desafíos de la gobernanza

El concepto de TI en la sombra, que se refiere al uso de sistemas de tecnología de la información sin la aprobación explícita del departamento de TI de la empresa, es un fenómeno ampliamente estudiado en la literatura sobre gestión de TI desde la década de 1980 (SILIC; BACK, 2014). Con la proliferación de plataformas públicas de IA, ha surgido una manifestación contemporánea y particularmente peligrosa de este fenómeno: la IA en la

[10]La norma ISO/IEC 42001:2023 es la primera norma internacional de sistemas de gestión dedicada específicamente a la inteligencia artificial. Establece requisitos y proporciona orientación a las organizaciones que desarrollan, suministran o utilizan sistemas de IA de forma responsable, abordando aspectos como la gobernanza, la gestión de riesgos, el impacto en los derechos humanos, la transparencia y la mejora continua.

[11]Una arquitectura de confianza cero (ZTA), según la define NIST SP 800-207, se basa en tres principios fundamentales: (1) todas las fuentes de datos y los servicios informáticos se consideran recursos; (2) todos (3) la comunicación está protegida independientemente de la ubicación en la red; (4) el acceso a los recursos individuales se concede por sesión, con verificación continua de identidad y contexto.

[12]Un agente de seguridad de acceso a la nube (CASB) es una solución de seguridad que se sitúa entre los usuarios empresariales y los proveedores de servicios en la nube, aplicando políticas de seguridad al tráfico de datos en tiempo real. Las funciones clave de un CASB incluyen la visibilidad del uso de los servicios en la nube, el control de acceso, la protección contra amenazas y el cumplimiento normativo.

Año VII, vol. 1, 2026 | Envío: 3 de julio de 2026 | Aceptación: 6 de julio de 2026 | Publicación: 9 de julio de 2026

sombra, caracterizada por el uso de herramientas de IA por parte de los empleados sin el conocimiento, la aprobación ni el control de la organización.

La IA en la sombra plantea un desafío de gobernanza único en comparación con la TI en la sombra tradicional. En primer lugar, la barrera de entrada es prácticamente inexistente: existen herramientas de IA de alto rendimiento disponibles de forma gratuita o a un coste mínimo, accesibles desde cualquier dispositivo con conexión a internet, sin necesidad de instalación de software ni configuración de red. En segundo lugar, las consecuencias potenciales son inmediatas y posiblemente irreversibles: los datos introducidos en plataformas de IA externas pueden utilizarse para reentrenar modelos, lo que hace que su posterior recuperación sea prácticamente imposible.

investigación realizada por Cyberhaven (2023) analizó el comportamiento de más de 1,6 millones de empleados y descubrió que el 11 % de los datos introducidos en las herramientas de IAG eran datos corporativos sensibles , en particular código fuente, datos financieros, estrategias comerciales y comunicaciones internas. Un estudio de LayerX Security (2023) reveló que solo el 8 % de las organizaciones encuestadas contaban con controles efectivos para supervisar y controlar el uso de IAG por parte de sus empleados.

Los vectores de riesgo de la IA en la sombra se pueden categorizar en tres dimensiones: (a) riesgo de confidencialidad, debido a la exposición de información sensible y propiedad intelectual a proveedores externos sin una base contractual adecuada; (b) riesgo legal, debido al potencial de violación de la LGPD (Ley General de Protección de Datos de Brasil) y las cláusulas contractuales de confidencialidad con clientes y socios; y (c) riesgo de integridad y confiabilidad, debido a la toma de decisiones basada en resultados de modelos no validados, potencialmente inexactos o sesgados.

Desde una perspectiva de gobernanza, abordar la IA en la sombra requiere un enfoque preventivo, de detección y correctivo simultáneo. Las medidas preventivas incluyen establecer políticas claras para el uso aceptable de herramientas de IA, comunicadas y capacitadas a todos los empleados; poner a disposición alternativas corporativas aprobadas; y crear procesos ágiles para evaluar y aprobar nuevas herramientas. Las medidas de detección incluyen monitorear el tráfico de red para identificar el acceso a plataformas de IA no autorizadas. Las medidas correctivas implican la respuesta a incidentes, la evaluación de daños y la adaptación de procesos y controles.

Es fundamental que los empleados no perciban el enfoque de gobernanza de la IA en la sombra como meramente restrictivo o punitivo. La literatura sobre gestión del cambio y comportamiento organizacional demuestra que las prohibiciones absolutas, sin alternativas

RCMOS – Revista Científica Multidisciplinar O Saber.
ISSN: 2675-9128. São Paulo - SP.
Año VII, vol. 1, 2026 | Envío: 3 de julio de 2026 | Aceptación: 6 de julio de 2026 | Publicación: 9 de julio de 2026

adecuadas, tienden a aumentar la resistencia de los empleados y a fomentar comportamientos clandestinos más sofisticados (SILIC; BACK, 2014). Por lo tanto, una gobernanza eficaz de la IA en la sombra debe equilibrar el control y la habilitación.

DIÁLOGO ENTRE EL DERECHO Y LA TECNOLOGÍA

La relación entre Derecho y Tecnología se ha caracterizado históricamente por asimetrías temporales: la regulación jurídica suele llegar después de la consolidación de los fenómenos tecnológicos, lo que genera lagunas normativas que los agentes económicos aprovechan en detrimento de los individuos y la comunidad. En el contexto de la Gestión de la Innovación y la Tecnología (GIT), este desfase regulatorio resulta especialmente pronunciado.

El diálogo entre el Derecho y la Tecnología en el ámbito de la IA exige superar importantes barreras epistemológicas. Los abogados, en su formación tradicional, no están preparados para comprender los mecanismos técnicos de los sistemas de gestión del derecho ; a su vez, los ingenieros y los científicos de datos suelen desconocer los requisitos legales y las implicaciones de sus decisiones de diseño. Establecer puentes entre estos ámbitos del conocimiento es indispensable . esencial para el desarrollo de una gobernanza eficaz de la IAG.

En el marco regulatorio brasileño, la LGPD (Ley General de Protección de Datos de Brasil) constituye la principal base legal para abordar los riesgos del tratamiento automatizado de datos (TAD) para los activos informacionales de las organizaciones. El artículo 20 de la LGPD, que garantiza al titular de los datos el derecho a solicitar la revisión de las decisiones tomadas exclusivamente sobre la base del tratamiento automatizado de datos personales, representa un estándar particularmente relevante en el contexto del TAD aplicado a los procesos de toma de decisiones corporativas.

A nivel internacional, el Reglamento de Inteligencia Artificial de la Unión Europea (Ley de IA de la UE) [13], aprobado en 2024, representa el primer marco regulatorio integral específicamente dirigido a los sistemas de IA. Si bien no es vinculante para Brasil, la Ley de IA influye en la regulación global de la IA y sirve como referencia importante para el proceso legislativo brasileño (PARLAMENTO EUROPEO, 2024).

[13]La Ley de IA de la UE (Reglamento (UE) 2024/1689) clasifica los siguientes sistemas de IA como de riesgo inaceptable y prohibidos en la Unión Europea: sistemas de puntuación social por parte de las autoridades públicas; explotación de vulnerabilidades en grupos específicos; identificación biométrica remota en tiempo real en espacios públicos (con excepciones limitadas); manipulación subliminal; y sistemas de categorización biométrica basados en características sensibles.

RCMOS – Revista Científica Multidisciplinar O Saber.
ISSN: 2675-9128. São Paulo - SP.
Año VII, vol. 1, 2026 | Envío: 3 de julio de 2026 | Aceptación: 6 de julio de 2026 | Publicación: 9 de julio de 2026

La perspectiva interdisciplinaria entre Derecho y Tecnología revela que la protección jurídica de los datos en el contexto de la IAG no puede reducirse al cumplimiento formal de la LGPD (Ley General de Protección de Datos de Brasil). En este sentido, el principio de Privacidad El concepto de protección de datos desde el diseño [14], articulado por Ann Cavoukian (2009) e incorporado al marco regulatorio del RGPD europeo e, implícitamente, al artículo 46 de la LGPD, establece que la protección de datos debe incorporarse al diseño de los sistemas y procesos desde su concepción, y no tratarse como un elemento externo de cumplimiento.

Reflexiones filosóficas sobre el poder, los datos y la inteligencia artificial.

Un análisis de las repercusiones del abuso de la información en la privacidad y la seguridad de la información en el contexto empresarial no puede obviar un fundamento filosófico que sitúa el fenómeno dentro de su dimensión más amplia: las relaciones de poder que impregnan la producción, el control y la explotación del conocimiento y la información en la sociedad contemporánea.

Michel Foucault, en "Vigilar y castigar" (1975), describe el panóptico benthamiano [15]como una metáfora de la sociedad disciplinaria moderna: una arquitectura de vigilancia en la que la mirada del observador, incluso si está potencialmente ausente, induce en el observado una conciencia permanente de su visibilidad, produciendo un efecto de autocontrol y docilidad. La IA, con su capacidad para procesar y analizar enormes cantidades de datos de comportamiento, representa una forma de panoptismo radicalmente nueva y más sofisticada, que ya no se ubica en una torre de observación, sino que se distribuye ubicuamente por las infraestructuras digitales de las organizaciones y la sociedad (FOUCAULT, 1975).

El concepto foucaultiano de biopoder [16]—el poder sobre la vida, que actúa sobre los cuerpos individuales y las poblaciones— encuentra su expresión contemporánea en los sistemas

[14]principio de privacidad El programa Privacy by Design fue desarrollado por Ann Cavoukian , excomisionada de Información y Privacidad de Ontario (Canadá), y establece siete principios fundamentales: proactividad en lugar de reactividad; privacidad como configuración predeterminada; privacidad integrada en el diseño; funcionalidad completa (suma positiva, no suma cero); seguridad de extremo a extremo; visibilidad y transparencia; y respeto por la privacidad del usuario.

[15]El panóptico, diseñado por Jeremy Bentham en el siglo XVIII, es una estructura carcelaria circular con una torre de vigilancia central, desde donde el guardia puede observar a todos los prisioneros sin ser visto. Foucault utiliza esta metáfora para describir el funcionamiento de las instituciones disciplinarias modernas —prisiones, escuelas, hospitales, fábricas— en las que la mera conciencia de la posibilidad de ser observado es suficiente para generar autocontrol.

[16]El concepto de «biopoder» fue desarrollado por Foucault en el primer volumen de «Historia de la sexualidad» (1976) y en cursos del Collège de France. El biopoder se manifiesta de dos formas: la anatomopolítica del cuerpo humano (disciplina de los cuerpos individuales) y la biopolítica de la población (regulación de los procesos

algorítmicos que clasifican, segmentan y condicionan el comportamiento de consumidores, trabajadores y ciudadanos. Al inferir patrones de comportamiento, tendencias psicológicas y categorías de identidad a partir de datos aparentemente neutrales, el biopoder algorítmico ejerce una forma de biopoder que opera con una eficiencia y una escala sin precedentes en la historia de la humanidad.

Zygmunt Bauman , con su concepto de modernidad líquida, ofrece un marco fundamental para comprender la fragilidad de las estructuras sociales e institucionales ante la velocidad de las transformaciones tecnológicas. En "Modernidad líquida" (2001), Bauman describe una condición social marcada por la inestabilidad, la fugacidad de los vínculos y la disolución de las estructuras sólidas que antaño organizaban la vida social y política. En el contexto de la IAG, la liquidez se manifiesta en la rapidez con la que surgen y se difunden nuevos modelos y plataformas, dejando obsoletos los marcos regulatorios y las prácticas de gobernanza incluso antes de su consolidación (BAUMAN, 2001).

Byung-Chul Han analiza las paradojas de la hiperconectividad y la exposición voluntaria de datos. Según Han, la sociedad contemporánea se caracteriza por una compulsión hacia la transparencia y la visibilidad: individuos y organizaciones exponen voluntariamente información personal y estratégica en plataformas digitales, impulsados por la lógica de la performatividad y la optimización. Este impulso hacia la transparencia, desde la perspectiva de Han, no representa la emancipación prometida por la retórica tecnológica, sino más bien una nueva forma de control, más eficaz por ser internalizada y aparentemente voluntaria (HAN, 2017).

En «La era del capitalismo de vigilancia» (2019), [17]Shoshana Zuboff ofrece el análisis más directamente relevante para el contexto del capitalismo de vigilancia corporativo. Zuboff describe el capitalismo de vigilancia como una nueva lógica económica que transforma la experiencia humana en materia prima para productos y servicios conductuales. Para Zuboff, las grandes plataformas tecnológicas —y los proveedores de capitalismo de vigilancia encajan perfectamente en esta categoría— extraen datos de comportamiento de los usuarios sin una

biológicos de las poblaciones). La IA analítica representa una nueva tecnología del biopoder que opera a una escala y granularidad sin precedentes.

[17]Zuboff acuñó el término "capitalismo de vigilancia" para describir una nueva lógica económica surgida de las prácticas de Google, que descubrió que los datos de comportamiento de los usuarios —originalmente un subproducto de las búsquedas— podían utilizarse para predecir e influir en comportamientos futuros, creando un mercado altamente lucrativo para la publicidad conductual. Esta lógica fue adoptada posteriormente por Facebook y otras plataformas digitales.

Año VII, vol. 1, 2026 | Envío: 3 de julio de 2026 | Aceptación: 6 de julio de 2026 | Publicación: 9 de julio de 2026

compensación adecuada, los procesan para predecir e influir en comportamientos futuros y los monetizan en el «mercado de futuros conductuales» (ZUBOFF, 2021).

La articulación de estos marcos filosóficos revela que los desafíos de la inteligencia artificial (IA) para la privacidad y la seguridad de la información no son meramente técnicos o legales, sino fundamentalmente políticos y éticos. La cuestión de quién controla los datos, con qué fines y con qué consecuencias para la autonomía individual y colectiva es, en última instancia, una cuestión de poder, y su respuesta requiere no solo regulación y tecnología, sino también una profunda reflexión ética sobre los valores que deben guiar el desarrollo y el uso de la inteligencia artificial.

Propuestas para una gobernanza de la IA segura y responsable

A partir del análisis interdisciplinario desarrollado en las secciones anteriores, esta sección presenta una serie de propuestas para la creación de un modelo de gobernanza de IA seguro y responsable en el entorno corporativo. En términos de gobernanza corporativa, se propone la creación de una estructura de gobernanza de IA que integre a representantes de las áreas legal, de seguridad de la información, tecnología, recursos humanos, cumplimiento normativo y alta dirección. La norma ISO/IEC 42001:2023 [18]proporciona el marco estructural para dicha gobernanza.

En materia legal y de cumplimiento normativo, se propone realizar una Evaluación de Impacto en la Protección de Datos (EIPD) [19], conforme a lo dispuesto en el artículo 38 de la LGPD (Ley General de Protección de Datos de Brasil), para cada caso de uso significativo de IA que involucre datos personales. Además, los contratos con los proveedores de plataformas de IA deben incluir cláusulas específicas sobre el tratamiento de datos, la confidencialidad, la prohibición del uso de datos para el reentrenamiento de modelos, la notificación de incidentes y la responsabilidad civil.

---

[19]La Evaluación de Impacto en la Protección de Datos (EIPD), también conocida como Evaluación de Impacto en la Protección de Datos La Evaluación de Impacto en la Protección de Datos (EIPD), según lo estipulado en el artículo 38 de la LGPD (Ley General de Protección de Datos de Brasil), es obligatoria cuando el tratamiento de datos pueda suponer riesgos para las libertades civiles y los derechos fundamentales. La ANPD (Autoridad Nacional de Protección de Datos de Brasil) podrá determinar las circunstancias en las que la EIPD será obligatoria.

Año VII, vol. 1, 2026 | Envío: 3 de julio de 2026 | Aceptación: 6 de julio de 2026 | Publicación: 9 de julio de 2026

En términos tecnológicos, la propuesta contempla la implementación de una arquitectura de seguridad basada en los principios de Confianza Cero, con controles específicos para el uso de IAG: políticas DLP configuradas para identificar y bloquear la transmisión de datos confidenciales a plataformas IAG no aprobadas; soluciones CASB para supervisar y controlar el uso de servicios IAG basados en la nube; una gestión rigurosa de identidades y accesos (IAM) para las herramientas IAG corporativas aprobadas; e implementación preferencial de entornos IAG privados para casos de uso que involucren datos confidenciales.

En lo que respecta a la cultura organizacional y la gestión de personas, se propone desarrollar un programa continuo de sensibilización y capacitación sobre el uso responsable de las herramientas de IA, abordando los riesgos de seguridad, las obligaciones legales derivadas de la LGPD (Ley General de Protección de Datos de Brasil), las políticas corporativas de uso aceptable y las herramientas aprobadas disponibles para satisfacer las necesidades de los empleados.

Desde un punto de vista ético, se propone adoptar un conjunto de principios para el uso responsable de la IA en el entorno empresarial, en consonancia con las directrices éticas de la Unión Europea para una IA fiable (COMISIÓN EUROPEA, 2019) y los principios de la OCDE sobre IA (OCDE, 2019): supervisión y control humanos; transparencia y explicabilidad; equidad algorítmica y no discriminación; privacidad y seguridad de los datos; responsabilidad y rendición de cuentas; y beneficio social y sostenibilidad ambiental.

Por último, la propuesta fomenta la participación activa de las organizaciones en los procesos de desarrollo normativo, aportando su experiencia práctica al desarrollo de marcos regulatorios que protejan los derechos fundamentales y promuevan la innovación responsable.

METODOLOGÍA

Este artículo adopta una metodología de investigación bibliográfica, exploratoria y cualitativa, basada en una revisión de la literatura científica, la legislación aplicable, las normas técnicas y los documentos institucionales.

La investigación bibliográfica abarcó obras de referencia en los campos del Derecho, el Derecho Digital, la Seguridad de la Información, la Filosofía de la Tecnología y el Gobierno Corporativo, con especial énfasis en las publicaciones de los últimos diez años, sin perjuicio de las obras clásicas de referencia filosófica. Las bases de datos consultadas incluyeron el Portal de Publicaciones Periódicas de CAPES, Google Académico, la Biblioteca Digital Brasileña de

Año VII, vol. 1, 2026 | Envío: 3 de julio de 2026 | Aceptación: 6 de julio de 2026 | Publicación: 9 de julio de 2026

Tesis y Disertaciones (BDTD) y los repositorios institucionales de universidades nacionales e internacionales.

La revisión de la literatura técnica y normativa abarcó las principales normas y marcos internacionales relevantes: ISO/IEC 27001:2022, ISO/IEC 27002:2022, ISO/IEC 42001:2023, Marco de Ciberseguridad del NIST (2018), Marco de Gestión de Riesgos de IA del NIST (2023) y OWASP Top 10 para Aplicaciones LLM (2023). Se examinó en su totalidad la legislación brasileña y extranjera aplicable, con especial atención a la LGPD (Ley General de Protección de Datos de Brasil) y la Ley de IA de la UE .

El carácter exploratorio de la investigación se debió a la relativa novedad del fenómeno estudiado y a la escasez de estudios empíricos específicos sobre el uso corporativo de la IA en el contexto brasileño. El enfoque cualitativo permitió el análisis interpretativo y crítico de los datos recopilados, con énfasis en la creación de conexiones interdisciplinarias entre los campos del Derecho, la Tecnología y la Filosofía.

El procedimiento metodológico adoptado comprendió: (a) búsqueda y selección de la literatura pertinente; (b) análisis crítico de los textos seleccionados; (c) síntesis y articulación interdisciplinaria de los contenidos; (d) formulación de propuestas basadas en el análisis realizado. Las referencias bibliográficas se organizaron y formatearon según las normas ABNT NBR 6023:2018.

CONSIDERACIONES FINALES

El análisis desarrollado a lo largo de este artículo demuestra que la Inteligencia Artificial Generativa representa, simultáneamente, una de las herramientas de innovación más poderosas a disposición de las organizaciones contemporáneas y un vector de riesgo legal, tecnológico y organizacional de magnitud sin precedentes. La tensión entre estos dos polos, innovación y riesgo, no puede resolverse mediante la simple adopción tecnológica ni la regulación aislada, sino que requiere un enfoque verdaderamente integrado.

El problema de investigación planteado no admite una respuesta simple ni universal. Cada organización tiene un perfil de riesgo único, su propia cultura organizacional y un contexto regulatorio específico. No obstante, los principios y prácticas identificados en esta investigación ofrecen un marco coherente para desarrollar respuestas contextualizadas y eficaces.

El análisis jurídico ha demostrado que la LGPD (Ley General de Protección de Datos de Brasil), si bien no fue redactada específicamente para el contexto de IAG, proporciona un

conjunto de principios e instrumentos regulatorios suficientemente sólidos para regir el uso corporativo de estas herramientas, siempre que se interprete y aplique de forma creativa y contextualizada. La responsabilidad de las organizaciones no se limita al cumplimiento formal de los requisitos legales, sino que también incluye adoptar una postura proactiva en la protección de los derechos de los titulares de los datos.

El análisis tecnológico reveló que los riesgos de la gestión de identidades y accesos (IAG) para la seguridad de la información son reales y potencialmente devastadores, pero se pueden gestionar mediante controles técnicos y organizativos adecuados. La arquitectura de confianza cero, las soluciones DLP y CASB, la gestión rigurosa de identidades y accesos, y el despliegue de entornos IAG privados representan lo último en controles técnicos disponibles.

El análisis filosófico ha revelado que los desafíos de la inteligencia centrada en el ser humano (HCI) trascienden las dimensiones técnicas y legales, situándose en el ámbito de las relaciones de poder, la vigilancia, la mercantilización de la experiencia humana y la erosión de la autonomía individual y colectiva. Los marcos teóricos de Foucault, Bauman , Han y Zuboff ofrecen perspectivas críticas indispensables para comprender las implicaciones más profundas de la HCI.

La conclusión fundamental de esta investigación es que la protección eficaz contra los riesgos de la IAG no depende únicamente de la tecnología o la legislación de forma aislada, sino de la integración de personas, procesos, gobernanza, seguridad y cultura organizacional. Las organizaciones que desarrollen esta integración de manera consistente y sostenible no solo cumplirán con los requisitos legales y las mejores prácticas de seguridad, sino que también estarán mejor posicionadas para aprovechar el verdadero valor que la IAG ofrece, de forma responsable, ética y sostenible.

El camino que tenemos por delante exige un esfuerzo colectivo: de las organizaciones, para desarrollar programas de gobernanza sólidos para la inteligencia artificial generativa (IAG); del Poder Legislativo, para construir marcos regulatorios específicos y apropiados; de la Autoridad Nacional de Protección de Datos (ANPD), para regular y orientar a los agentes de procesamiento de datos; de la academia, para generar conocimiento interdisciplinario riguroso; y de la sociedad civil, para monitorear y defender los derechos fundamentales amenazados por el uso irresponsable de estas tecnologías. La inteligencia artificial generativa es, en última instancia, un reflejo de la inteligencia humana, y las decisiones que tomamos sobre cómo desarrollarla y usarla reflejan, para bien o para mal, los valores que como sociedad decidimos priorizar.

Año VII, vol. 1, 2026 | Envío: 3 de julio de 2026 | Aceptación: 6 de julio de 2026 | Publicación: 9 de julio de 2026

REFERENCIAS

AUTORIDAD NACIONAL DE PROTECCIÓN DE DATOS (ANPD). Guía sobre la transferencia internacional de datos personales. Brasilia: ANPD, 2023. Disponible en: https://www.gov.br/anpd. Consultado el 8 de junio de 2026.

BAUMAN, Zygmunt. Modernidad líquida. Traducido por Plínio Dentzien . Río de Janeiro: Zahar, 2001.

BOMMASANI, Rishi et al. Sobre las oportunidades y los riesgos de los modelos de fundación. Stanford CRFM, 2021. Disponible en: https://arxiv.org/abs/2108.07258. Consultado el 8 de junio de 2026.

BRASIL. Ley N° 13.709, de 14 de agosto de 2018. Ley General de Protección de Datos Personales (LGPD). Boletín Oficial . Gaceta Oficial de la Unión : Sección 1, Brasilia, DF, 15 de agosto de 2018.

BUSINESSINSIDER. Empleados de Samsung filtraron accidentalmente secretos comerciales mientras usaban ChatGPT. Business Insider, 2023. Disponible en: https://www.businessinsider.com/samsung-workers-leaked-trade-secrets-chatgpt-2023-4. Acceso 8 de junio de 2026.

CAVOUKIAN, Ann. Privacidad desde el diseño: Los 7 principios fundamentales. Ontario: Comisionado de Información y Privacidad de Ontario, 2009.

CYBERHAVEN. Estudio de Cyberhaven con un millón de usuarios: Cómo los empleados utilizan ChatGPT y otras herramientas de IA en el trabajo. San Francisco: Cyberhaven , 2023.

COMISIÓN EUROPEA. Directrices éticas para una IA fiable. Bruselas : Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial de la Comisión Europea, 2019.

PARLAMENTO EUROPEO. Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas sobre inteligencia artificial (Ley de Inteligencia Artificial). Diario de el Unión Europea , 2024.

FOUCAULT, Michel. Vigilar y castigar: El nacimiento de la prisión. Traducido por Raquel Ramalhete. Petrópolis : Vozes , 1987. (Edición original publicada) en 1975).

GARTNER. Principales tendencias tecnológicas estratégicas para 2024. Stamford: Gartner Research, 2023.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. Aprendizaje profundo. Cambridge: MIT Press, 2016.

HAN, Byung -Chul. La Sociedad de la Transparencia. Traducido por Enio Paulo Giachini . Petrópolis: Vozes, 2017. (Original publicado en 2012).

HAN, Byung -Chul. En el enjambre: Perspectivas sobre lo digital. Traducido por Lucas Machado. Petrópolis: Vozes, 2018. (Original publicado) en 2013).

ORGANIZACIÓN INTERNACIONAL DE NORMALIZACIÓN. ISO/IEC 27001:2022 – Seguridad de la información, ciberseguridad y protección de la privacidad – Sistemas de gestión de la seguridad de la información – Requisitos. Ginebra : ISO, 2022.

ORGANIZACIÓN INTERNACIONAL DE NORMALIZACIÓN. ISO/IEC 27002:2022 – Seguridad de la información, ciberseguridad y protección de la privacidad – Controles de seguridad de la información. Ginebra : ISO, 2022.

Año VII, vol. 1, 2026 | Envío: 3 de julio de 2026 | Aceptación: 6 de julio de 2026 | Publicación: 9 de julio de 2026

ORGANIZACIÓN INTERNACIONAL DE NORMALIZACIÓN. ISO/IEC 42001:2023 – Inteligencia artificial – Sistema de gestión. Ginebra : ISO, 2023.

LAYERX SECURITY. Informe anual de seguridad de navegadores 2023. Tel Aviv: LayerX Security, 2023.

MCKINSEY GLOBAL INSTITUTE. El potencial económico de la IA generativa: la próxima frontera de la productividad. Nueva York: McKinsey & Company, 2023.

MICROSOFT. Azure OpenAI Service: Datos, privacidad y seguridad para Azure OpenAI Service. Redmond: Microsoft, 2023. Disponible en: https://learn.microsoft.com/azure/ai-services/openai/. Consultado el [fecha] . 8 de junio de 2026.

INSTITUTO NACIONAL DE ESTÁNDARES Y TECNOLOGÍA (NIST). Marco de gestión de riesgos de inteligencia artificial (AI RMF 1.0). NIST AI 100-1. Gaithersburg: NIST, 2023.

INSTITUTO NACIONAL DE ESTÁNDARES Y TECNOLOGÍA (NIST). Marco para la mejora de la ciberseguridad de la infraestructura crítica, versión 1.1. Gaithersburg: NIST, 2018.

INSTITUTO NACIONAL DE ESTÁNDARES Y TECNOLOGÍA (NIST). Arquitectura de confianza cero. Publicación especial NIST 800-207. Gaithersburg: NIST, 2020.

OCDE. Recomendación del Consejo sobre Inteligencia Artificial. OECD/LEGAL/0449. París: OCDE, 2019.

OPENAI. Informe técnico de GPT-4. San Francisco: OpenAI, 2023. Disponible en: https://arxiv.org/abs/2303.08774. Consultado el 8 de junio de 2026.

OUYANG, Long et al. Entrenamiento de modelos de lenguaje para seguir instrucciones con retroalimentación humana. En: AVANCES EN SISTEMAS DE PROCESAMIENTO DE INFORMACIÓN NEURAL, 35, 2022. Actas [...]. NeurIPS , 2022.

Fundación OWASP. Los 10 principales problemas de OWASP para aplicaciones de modelos de lenguaje a gran escala. Versión 1.1. Wilmington: OWASP, 2023. Disponible en: https://owasp.org/www-project-top-10-for-large-language-model-applications/. Consultado el 8 de junio de 2026.

SILIC, Mario; BACK, Andrea. Shadow IT: Una visión desde detrás del telón. Computers & Security, vol. 45, págs. 274-283, 2014.

THE ECONOMIST. El recurso más valioso del mundo ya no es el petróleo, sino los datos. The Economist, Londres, 6 de mayo de 2017.

VASWANI, Ashish et al. La atención es todo lo que necesitas. En: AVANCES EN SISTEMAS DE PROCESAMIENTO DE INFORMACIÓN NEURAL, 30, 2017. Actas [...]. NeurIPS , 2017.

FORO ECONÓMICO MUNDIAL. Flujo libre de datos con confianza: superando las barreras a los flujos de datos transfronterizos. Cologny /Ginebra: Foro Económico Mundial , 2020.

ZUBOFF, Shoshana . La era del capitalismo de vigilancia: La lucha por un futuro humano en la nueva frontera del poder. Traducido por George Schlesinger. Río de Janeiro: Intrínseca, 2021. (Publicado originalmente en 2019).