EST. 2021 EMC
EDITORIAL MAR CARIBE

PROTECCIÓN DE SECRETOS EMPRESARIALES VS. TRANSPARENCIA ALGORÍTMICA

ESCRITO POR:

SANTIAGO GONZALES MESIA
ROBERTO SEGUNDO TEJADA RODRIGUEZ
JOSÉ LUIS CASTRO ULLILEN
VIVIANA INÉS VELLÓN FLORES DE SOLANO
TIMOTEO SOLANO ARMAS
CARLOS MÁXIMO GONZÁLES AÑORGA

ISBN: 978-9915-698-63-2

Protección de secretos empresariales vs. transparencia algorítmica

Gonzales Mesia, Santiago; Tejada Rodriguez, Roberto Segundo; Castro Ullilen, José Luis; Vellón Flores de Solano, Viviana Inés; Solano Armas, Timoteo; Gonzáles Añorga, Carlos Máximo

Primera edición (1.ª ed.): febrero, 2026

Editado por:

Editorial Mar Caribe ®

www.editorialmarcaribe.es

Av. Gral. Flores 547, 70000 Col. del Sacramento, Departamento de Colonia, Uruguay.

Diseño de carátula e ilustraciones: *Luisa Fernanda Lugo Rojas*

Libro electrónico disponible en:

https://editorialmarcaribe.es/ark:/10951/isbn.9789915698632

Formato: Electrónico

ISBN: 978-9915-698-63-2

ARK: ark:/10951/isbn.9789915698632

Editorial Mar Caribe (OASPA): Como miembro de la Open Access Scholarly Publishing Association, apoyamos el acceso abierto de acuerdo con el código de conducta, la transparencia y las mejores prácticas de OASPA para la publicación de libros académicos y de investigación. Estamos comprometidos con los más altos estándares editoriales en ética y deontología, bajo la premisa de «Ciencia Abierta en América Latina y el Caribe»

OASPA

Editorial Mar Caribe, firmante Nº 795 de 12.08.2024 de la Declaración de Berlín

"... Nos sentimos obligados a abordar los retos de Internet como un medio funcional emergente para la distribución del conocimiento. Obviamente, estos avances pueden modificar significativamente la naturaleza de la publicación científica, así como el actual sistema de garantía de calidad...." (Max Planck Society, ed. 2003, pp. 152-153).

Editorial Mar Caribe

Protección de secretos empresariales vs. transparencia algorítmica

Colonia, Uruguay

2026

Protección de secretos empresariales vs. transparencia algorítmica

Índice

Introducción

Actualmente, la forma en que tomamos decisiones ha cambiado radicalmente. Hemos pasado de una administración basada en la burocracia humana, sometida a revisión y control directo, a una algocracia en la que los sistemas de Inteligencia Artificial (IA) y Machine Learning se encargan de tareas que van desde otorgar créditos bancarios hasta evaluar riesgos en procesos penales. En este contexto, el algoritmo se convierte en el nuevo eje central del poder económico y social.

No obstante, esta eficiencia tecnológica se acompaña de un fenómeno preocupante: la opacidad. La complejidad de los modelos de aprendizaje profundo ha dado lugar al denominado problema de la caja negra, en el que incluso los desarrolladores tienen dificultades para explicar el razonamiento que llevó a un resultado concreto.

Desde la perspectiva del Derecho de la Propiedad Intelectual e Industrial, el algoritmo se considera un activo intangible de valor incalculable. La Directiva (UE) 2016/943 y leyes nacionales, como la Ley 1/2019 de Secretos Empresariales en España, han establecido un marco de protección sólido para la información confidencial que tiene valor comercial y está protegida mediante medidas razonables de confidencialidad.

Para las empresas tecnológicas, mantener ciertos inventos en secreto suele ser más ventajoso que solicitar una patente, ya que evita la divulgación del invento y no conlleva una fecha de expiración fija. El secreto empresarial no solo es una protección legal, sino también un estímulo económico que impulsa la inversión en investigación y desarrollo en un mercado global

altamente competitivo.

Frente al derecho de las empresas a proteger sus activos, se alza un clamor jurídico y ético por la transparencia de los algoritmos. Este no debe entenderse solo como el acceso al código fuente —que generalmente es difícil de entender para el ciudadano promedio—, sino como el derecho a obtener una explicación clara y significativa sobre cómo funciona la lógica de la decisión, cuáles son sus criterios de ponderación y qué datos de entrenamiento se usaron.

La falta de transparencia implica riesgos sistémicos: la persistencia de sesgos en los algoritmos, la discriminación automática y la violación de derechos fundamentales, como el derecho a la defensa y al debido proceso. Cuando un algoritmo determina el destino de una persona, el

¿Es posible lograr una coexistencia equilibrada? La idea central de esta investigación es la aparente imposibilidad de conciliar ambos conceptos. ¿Puede una empresa mantener la transparencia sin perder su ventaja competitiva? ¿Es factible que el Estado proteja los derechos fundamentales sin desalentar la innovación tecnológica?

Este libro busca analizar si el marco jurídico actual es suficiente para abordar estas tensiones o si, en cambio, es necesaria una Transparencia 4.0. Esta transparencia no se lograría mediante la divulgación pública del algoritmo, sino a través de mecanismos de auditoría independiente, entornos regulatorios controlados y la adopción de una explicabilidad por diseño (explainability by design).

La expansión de los sistemas de Inteligencia Artificial (IA) en decisiones críticas —como otorgar créditos, seleccionar personal o administrar la justicia penal— ha provocado una tensión jurídica sin precedentes. La razón de este

libro se fundamenta en los siguientes pilares:

- *El dilema de la caja negra y el debido proceso:*

Los algoritmos a menudo funcionan como cajas negras cuya lógica interna no es accesible. Cuando estas decisiones afectan derechos fundamentales, se requiere mayor transparencia. Sin embargo, las empresas que los desarrollan sostienen que divulgar el funcionamiento detallado de sus modelos pondría en riesgo sus secretos comerciales, lo que comprometería su ventaja competitiva y su motivación para innovar.

- *Vacíos en el marco regulatorio actual:*

A pesar de que normativas recientes como el AI Act de la Unión Europea han sido aprobadas, aún existe incertidumbre sobre cómo equilibrar el derecho a la explicación (transparencia) con la protección de la propiedad intelectual. La investigación resulta necesaria para proponer criterios de ponderación que prevengan que el secreto empresarial se convierta en una excusa para la opacidad o la discriminación por parte de los algoritmos.

- *Impacto en la confianza social y la innovación:*

La falta de transparencia socava la confianza de los consumidores en las tecnologías emergentes, mientras que una transparencia total y sin protección puede desalentar la inversión privada en I+D. Este libro busca encontrar el equilibrio clave para mantener un ecosistema digital sostenible.

Ahora se invita al lector a reflexionar sobre cómo crear una sociedad digital que sea tanto innovadora como justa: una comunidad donde la tecnología impulse el progreso sin ocultarse tras lo misterioso, con base en el objetivo de investigación: proporcionar una hoja de ruta doctrinal y práctica para legisladores, jueces y desarrolladores que permita la rendición de cuentas algorítmica sin desproteger los activos intangibles esenciales de la economía

digital.

Capítulo 1

El conflicto normativo entre la protección de secretos empresariales y la transparencia algorítmica

La transición global hacia una economía de datos y el despliegue masivo de sistemas de inteligencia artificial (IA) han precipitado una colisión jurídica fundamental entre dos pilares de la sociedad moderna: la protección de los activos intangibles mediante el secreto empresarial y el derecho a la transparencia, la explicabilidad y la rendición de cuentas algorítmica. Esta tensión no es simplemente un debate técnico sobre el acceso al código fuente, sino una disputa profunda sobre la distribución del poder en la era digital, la protección de los derechos humanos y la capacidad de las instituciones democráticas para supervisar decisiones automatizadas que afectan la vida, la libertad y el bienestar de los ciudadanos.

Mientras las empresas argumentan que la opacidad es necesaria para proteger la innovación y la ventaja competitiva, la sociedad civil y los reguladores sostienen que la caja negra algorítmica facilita sesgos, discriminación y una arbitrariedad tecnológica incompatible con el Estado de derecho.

La naturaleza jurídica del algoritmo y la primacía del secreto empresarial

En el ecosistema de la inteligencia artificial, el algoritmo constituye el núcleo de la ventaja competitiva. Su protección no se limita a la mera expresión literal del código, tradicionalmente protegida por el derecho de autor, sino que abarca la lógica subyacente, los pesos de las variables, las metodologías de entrenamiento y los conjuntos de datos que permiten al sistema generar predicciones o recomendaciones precisas. Las empresas han optado mayoritariamente por el régimen de secretos empresariales (trade secrets) frente al sistema de patentes por razones estratégicas de peso: flexibilidad, ausencia de registros públicos, duración potencialmente ilimitada y un alcance que protege no solo la invención final, sino también el saber hacer (know-how) y los procesos intermedios (De Noyette et al., 2025).

El marco legal que sustenta esta protección, ejemplificado por la Directiva (UE) 2016/943 y por leyes como la Ley 1/2019 en España o la Ley 19.39 en Chile, exige que la información sea secreta, tenga valor comercial derivado de ese secreto y sea objeto de medidas razonables para mantenerla oculta. Si el valor de un modelo de calificación crediticia o de un sistema de selección de personal reside en la exclusividad de su lógica interna, cualquier mandato de apertura del código o de explicación detallada de sus variables se percibe como una amenaza directa a la viabilidad económica del desarrollador (Viveros, 2015). Sin embargo, esta protección no es absoluta y debe ceder ante intereses públicos superiores, como la prevención de la discriminación y la protección de los derechos fundamentales.

El imperativo de la transparencia algorítmica: Fundamentos y dimensiones

Frente al secretismo industrial, la transparencia algorítmica emerge no solo como un principio ético, sino también como una necesidad jurídica para garantizar la legalidad de los actos administrativos y privados. La opacidad de los sistemas de toma de decisiones automatizados (SDA) plantea riesgos sistémicos, que van desde la perpetuación de sesgos históricos hasta la creación de nuevas formas de exclusión digital. La transparencia algorítmica debe entenderse de manera multidimensional, abarcando desde la simple notificación de que se interactúa con una IA hasta la explicabilidad profunda de la lógica de una decisión individual (Araya, 2021).

La explicabilidad como garantía de derechos

La explicabilidad (explainability) es la dimensión de la transparencia que permite a un ser humano comprender las razones de un resultado algorítmico. El uso generalizado de la inteligencia artificial está impactando a la sociedad de formas aún poco claras, lo que se traduce en efectos negativos, como el aumento de las desigualdades sociales debido a decisiones algorítmicas (Ozmen et al., 2023). En sectores sensibles como la justicia penal o el empleo, no basta con saber que se utilizó un algoritmo; es necesario entender qué factores determinaron una decisión negativa.Esto ha dado lugar al desarrollo de la Inteligencia Artificial Explicable (XAI), que utiliza técnicas para ofrecer interpretaciones legibles por humanos sin necesidad de revelar la totalidad del código fuente propietario.

El problema de la caja negra (Black Box)

Muchos modelos de aprendizaje profundo (deep learning) operan mediante redes neuronales con millones de parámetros que resultan inescrutables incluso para sus propios creadores. Esta opacidad técnica se suma a la opacidad jurídica del secreto empresarial, creando un vacío de responsabilidad en el que nadie —ni el desarrollador, ni el usuario, ni el afectado— puede explicar plenamente por qué ocurrió un daño o se produjo un sesgo. La auditoría algorítmica surge entonces como el mecanismo para romper este bloqueo, permitiendo evaluaciones externas de los riesgos de precisión, de justicia y de seguridad.

El reglamento de IA de la Unión Europea (RIA) y la transparencia cualificada

La Unión Europea ha establecido el estándar global de regulación mediante el Reglamento de Inteligencia Artificial (RIA), que introduce un enfoque basado en el riesgo para equilibrar la innovación con la seguridad. El RIA no exige una transparencia absoluta para todos los sistemas, sino que impone obligaciones proporcionales al impacto potencial de cada sistema sobre los derechos fundamentales (véase la Tabla 1).

La Ley de IA busca que los europeos confíen en la tecnología. Aunque la mayoría de los sistemas de IA presentan poco riesgo, algunos pueden causar problemas difíciles de detectar, como decisiones injustas en contrataciones o solicitudes de beneficios, ya que no siempre es posible entender cómo llegan a sus conclusiones.

Tabla 1: Reglamento de IA de la UE (RIA) y transparencia cualificada

Nivel de riesgo	Ejemplos de aplicación	Obligaciones de transparencia
Inaceptable	Social scoring, vigilancia biométrica masiva.	Prohibición total de uso.
Alto Riesgo	Salud, justicia, empleo, servicios esenciales.	Documentación técnica, registro, supervisión humana.
Limitado	Chatbots, generación de contenido sintético.	Notificación de uso y marcado de contenido (watermarking).
Mínimo	Filtros de spam, videojuegos.	Sin obligaciones específicas: códigos de conducta voluntarios.

La tensión en los sistemas de alto riesgo

Es en la categoría de alto riesgo donde el conflicto con el secreto empresarial es más acuciante. Los proveedores deben elaborar documentación

técnica detallada que cubra la lógica general, las metodologías de entrenamiento y los conjuntos de datos. El Artículo 78 del RIA estipula que, aunque la información confidencial debe protegerse, ello no debe impedir la aplicación efectiva del reglamento. Esto introduce el concepto de transparencia cualificada: la información fluye a los reguladores y auditores bajo estrictos deberes de confidencialidad, pero no necesariamente al público en general, lo que protege el valor comercial del activo.

El régimen para modelos de IA de propósito general (GPAI)

El RIA también aborda los modelos de IA de propósito general, como los que sustentan a ChatGPT. Estos modelos deben proporcionar resúmenes de los datos de entrenamiento para permitir a los titulares de derechos de autor ejercer sus facultades de exclusión (opt-out). Aquí, la transparencia sirve no solo para la rendición de cuentas social, sino también para el equilibrio del mercado de la propiedad intelectual (Morales, 2021).

La situación en América Latina: convergencia y particularidades regionales

América Latina ha iniciado un proceso de efecto Bruselas, adaptando los principios del RIA europeo a sus marcos legislativos nacionales, con un énfasis creciente en la protección de los derechos humanos frente a los sistemas predictivos en los servicios públicos (Beliz, 2025).

Chile: Liderazgo en gobernanza basada en riesgo

Chile ha avanzado significativamente con un proyecto de ley (Boletín 16821-19) que clasifica los sistemas de IA de manera casi idéntica a la del modelo europeo. El marco chileno subraya que los ciudadanos tienen derecho a saber cuándo una decisión que les afecta (como la denegación de un crédito o de un subsidio) ha sido tomada por una máquina y a solicitar explicaciones sobre la lógica aplicada. No obstante, el debate en el Senado ha puesto de relieve preocupaciones sobre la capacidad de la futura Agencia de Protección de Datos Personales para fiscalizar estos sistemas sin asfixiar la innovación local.

Brasil: el Proyecto de Ley 2338/2023 y el Sistema Nacional (SIA)

Brasil ha optado por un enfoque que integra la IA en su robusto marco de protección de datos (LGPD). El proyecto de ley 2338/2023 establece el derecho a la explicación como pilar fundamental y prevé la creación de un Sistema Nacional de Regulación y Gobernanza de la IA (SIA). A diferencia de otros modelos, el brasileño enfatiza la necesidad de auditorías periódicas en sistemas de alto riesgo, permitiendo excepciones al secreto comercial únicamente cuando sea necesario verificar la ausencia de sesgos discriminatorios (Loayza, 2025).

México y Colombia: enfoque en transparencia pública

En general, el avance se ha dado principalmente en el ámbito del derecho administrativo. Se han emitido directrices para que los algoritmos

utilizados por el Estado en la gestión de los servicios sociales y tributarios sean auditables por los organismos garantes de la transparencia (Viveros, 2015). El informe Transparencia Algorítmica 2025 para la región subraya que el secreto industrial no puede invocarse por parte de empresas privadas que contratan con el Estado para ocultar la lógica de los sistemas que determinan el acceso a derechos fundamentales.

Jurisprudencia crítica: Los tribunales como árbitros de la opacidad

La lucha por la transparencia algorítmica ha tenido sus capítulos más dramáticos en las cortes superiores, donde se han ponderado el debido proceso y la propiedad intelectual.

El caso State v. Loomis y la lección de COMPAS

En Estados Unidos, el caso *Loomis v. Wisconsin* (2016-2017) desafió el uso de COMPAS, un algoritmo propietario utilizado para evaluar el riesgo de reincidencia en sentencias penales. El acusado, Eric Loomis, argumentó que su derecho al debido proceso fue vulnerado al ser sentenciado con base en una puntuación cuya metodología era un secreto comercial. Aunque el tribunal no prohibió el uso del algoritmo, impuso advertencias cruciales: el sistema no puede ser el único factor determinante y los jueces deben ser informados de que la naturaleza propietaria impide una validación científica externa completa por parte de la defensa.

Este caso evidenció que el secreto comercial puede ocultar sesgos

raciales y de género, en el que el algoritmo actúa como una patada en la espalda institucional que legitima prejuicios históricos bajo una pátina de objetividad técnica.

El hito del caso Bosco en España

En septiembre de 2025, el Tribunal Supremo de España dictó una sentencia histórica en el Caso Bosco. La administración se había negado a entregar el código fuente del algoritmo utilizado para la concesión del bono social eléctrico, alegando derechos de propiedad intelectual del desarrollador y riesgos para la seguridad pública. El tribunal falló a favor de la transparencia y determinó que, cuando un algoritmo se utiliza para adoptar decisiones administrativas con efectos jurídicos sobre los ciudadanos, debe considerarse un documento administrativo sujeto al derecho de acceso a la información pública. La sentencia aclaró que la protección de la propiedad intelectual no es un cheque en blanco para la opacidad del Estado y que prevalece el interés público en conocer los criterios de reparto de fondos públicos.

Gestión algorítmica del trabajo: de la Ley Rider a la subordinación invisible

El ámbito laboral es uno de los frentes más activos en la demanda de transparencia. Los algoritmos no solo seleccionan candidatos (como el sistema ATS o las herramientas de Gild), sino que también gestionan el día a día de millones de trabajadores de plataformas, decidiendo las asignaciones de tareas, las evaluaciones de desempeño y los despidos (Ladosky y López, 2025).

La transparencia como derecho laboral

España fue pionera con la Ley Rider (2021), que introdujo el derecho de los sindicatos a conocer los parámetros de los algoritmos que afectan las condiciones de trabajo. Esta medida busca reequilibrar la asimetría de información entre la empresa y el trabajador. Sin transparencia, el trabajador se enfrenta a una gestión por algoritmo en la que las penalizaciones por baja productividad o ausencias (incluso por enfermedad, como en el caso de Deliveroo en Italia) se aplican de forma automática y opaca, lo que impide el ejercicio de los derechos de defensa (Ladosky y López, 2025).

El riesgo de la perfilación y el capital social

Empresas como Gild utilizan algoritmos para medir el capital social y predecir cuándo un empleado cambiará de trabajo. Estas herramientas evalúan variables, como la estabilidad emocional, a partir de la huella digital. La transparencia aquí es vital para evitar que el algoritmo castigue a individuos basándose en correlaciones espurias o discriminatorias que no guardan relación con su competencia profesional.

Metodologías de auditoría y soluciones técnicas: el enfoque de la caja negra

Para resolver el impasse entre secreto y transparencia, han surgido metodologías de auditoría que permiten el escrutinio sin la exposición total de los activos. El Banco Interamericano de Desarrollo (BID), a través de su iniciativa fAIr LAC, ha propuesto un marco de auditoría de caja negra que se

ha convertido en una referencia para la región.

El proceso de auditoría algorítmica

Una auditoría algorítmica no consiste simplemente en leer código, sino en un estudio sistemático e independiente que evalúa el ciclo de vida completo del sistema (véase la Tabla 2).

- **Fase de concepción y diseño:** evaluación de la necesidad de la IA y de los riesgos éticos iniciales.
- **Preparación de datos:** verificación de la representatividad y la calidad de los datos para evitar sesgos durante el entrenamiento.
- **Desarrollo y modelado:** pruebas de métricas de justicia (fairness) y de precisión.
- **Despliegue y monitoreo:** evaluación continua para detectar la deriva del modelo (drift) y fallos en entornos reales.

Tabla 2: Despliegue y monitoreo de auditoría algorítmica

Pilar de la Auditoría	**Objetivo Técnico**	**Requisito de información**
Gobernanza	Definir responsabilidades humanas.	Organigrama y procesos de decisión.
Justicia algorítmica	Detectar impacto dispar en grupos.	Datos de entrada y salida anonimizados.

Explicabilidad	Entender el porqué de un resultado.	Mapas de características e importancia de variables.
Privacidad	Cumplimiento de LGPD/RGPD.	Inventario de datos personales y cifrado.

Técnicas de XAI (Explainable AI)

Las herramientas de XAI actúan como traductores entre la complejidad matemática del modelo y la necesidad jurídica de explicación.

- **SHAP (SHapley Additive exPlanations):** Permite cuantificar con precisión cuánto contribuyó cada factor (edad, ingresos, historial) a una decisión de crédito específica.
- **LIME:** Aproxima el modelo complejo localmente mediante uno más sencillo para explicar predicciones individuales.
- **Gráficos de Dependencia Parcial (PDP):** Muestran el efecto marginal de una o dos variables sobre el resultado previsto.

Estas técnicas permiten cumplir con el derecho a la explicación exigido por el RGPD y por las nuevas leyes latinoamericanas, sin que la empresa tenga que entregar su código fuente a la competencia.

El conflicto entre los secretos empresariales y la transparencia algorítmica está evolucionando de una confrontación binaria a un modelo de transparencia proporcional y cualificada (Azuaje y Finol, 2020). La protección

de la propiedad intelectual ya no se acepta como un escudo absoluto contra la rendición de cuentas, especialmente cuando la IA se despliega en ámbitos que afectan la dignidad y los derechos fundamentales.

Para el año 2026, se espera que la plena implementación del Reglamento de IA de la UE y la aprobación de las leyes en Chile y Brasil consoliden un estándar global en el que la transparencia sea la regla y la confidencialidad, la excepción estrictamente necesaria. Las empresas que integren la explicabilidad desde el diseño y se sometan a auditorías independientes no solo cumplirán con la ley, sino que también obtendrán una ventaja competitiva basada en la confianza y la legitimidad social. En última instancia, la transparencia algorítmica no debe verse como un obstáculo para la innovación, sino como la infraestructura ética necesaria para que la inteligencia artificial sea verdaderamente confiable y beneficiosa para la humanidad (Araya, 2021).

Capítulo 2
Gobernanza de la inteligencia artificial: taxonomía institucional

La evolución de la gobernanza de la inteligencia artificial ha pasado de ser una preocupación periférica a convertirse en el eje estructurante de las políticas públicas y las estrategias corporativas a nivel global. La velocidad vertiginosa con la que avanzan las capacidades computacionales y algorítmicas ha forzado a los legisladores, organismos internacionales y entidades de la sociedad civil a estructurar un entramado institucional capaz de contener los riesgos inherentes sin sofocar la innovación tecnológica (Cihon et al., 2021).

Este ecosistema en rápida expansión no opera de manera homogénea ni centralizada, sino que se manifiesta a través de un complejo de regímenes fragmentado que abarca desde directrices éticas voluntarias hasta tratados internacionales vinculantes y leyes federales estrictas (Oncioiu y Bularca, 2025). El análisis de las respuestas regulatorias revela que el principal desafío radica en cerrar la brecha estructural entre la rapidez de la innovación técnica y los plazos de maduración prolongados de los procesos legislativos tradicionales.

Para comprender la magnitud y el alcance de este ecosistema, diversos análisis clasifican las más de 140 instituciones involucradas en la materia mediante una arquitectura de gobernanza de referencia, dividida en cuatro

capas independientes pero profundamente interrelacionadas. Esta taxonomía permite a los analistas de políticas públicas y a los oficiales de cumplimiento navegar por un océano de normativas que, de otro modo, resultaría inmanejable debido a la superposición de competencias y a la ambigüedad semántica que rodea al término «regulación de la inteligencia artificial».

La Tabla 3 detalla la distribución de estas capas institucionales y proporciona una base estructurada para evaluar el origen y la naturaleza de las obligaciones normativas que enfrentan las organizaciones que desarrollan o implementan sistemas algorítmicos.

Tabla 3: Naturaleza de las obligaciones normativas para las organizaciones

Capa de gobernanza	**Instituciones representativas**	**Ámbito de acción y enfoque**
Capa 1: Gobernanza Global	Naciones Unidas, enviado de tecnología de la ONU, UNESCO, OCDE, GPAI	Coordinación de políticas globales, establecimiento de principios éticos transversales y defensa de los derechos humanos.
Capa 2: Estándares Técnicos	ISO/IEC, IEEE, NIST, UIT	Desarrollo de marcos de gestión de riesgos, protocolos de interoperabilidad y especificaciones técnicas de seguridad.
Capa 3: Autoridades Reguladoras	Autoridades del Reglamento de IA de la UE, FTC, OPC, agencias de protección de datos	Fiscalización del cumplimiento normativo, imposición de sanciones administrativas y protección de los

		derechos de los consumidores.
Capa 4: Supervisión Ética y de Investigación	CIOMS, WMA, AEA, Comités de Ética en Investigación	Vigilancia de la ética aplicada en entornos específicos como la salud, la investigación clínica y el desarrollo académico.

El análisis de esta distribución institucional sugiere una tendencia hacia la consolidación de bloques de poder regulatorio, en la que las directrices emanadas de la primera capa sirven como justificación normativa para las reglas vinculantes de la tercera capa, y se emplean los estándares técnicos de la segunda capa como métrica de verificación empírica (Ruiz, 2023). No obstante, la proliferación de directrices superpuestas entraña el riesgo latente de generar una fatiga del derecho blando, en la que la multiplicación de códigos de conducta voluntarios diluye la rendición de cuentas en lugar de fortalecerla.

El tránsito del derecho blando al derecho duro: del consenso ético a la coerción legal

El análisis histórico de la gobernanza tecnológica demuestra que la fase inicial de la regulación de la inteligencia artificial estuvo dominada casi exclusivamente por el derecho blando. Esta categoría abarca principios, directrices y marcos de recomendaciones que, al carecer de fuerza vinculante directa, ejercen influencia sobre el comportamiento de los actores mediante el consenso, la presión de grupo y los incentivos reputacionales. El derecho

blando demostró ser sumamente útil en las primeras etapas de despliegue de los modelos de aprendizaje automático, pues sirvió como un laboratorio viviente. En este espacio de experimentación, los legisladores pudieron probar conceptos complejos —tales como las evaluaciones de impacto, la explicabilidad algorítmica y la auditoría continua— sin congelar la innovación tecnológica bajo leyes rígidas difíciles de actualizar (Caiza et al., 2024).

La Organización para la Cooperación y el Desarrollo Económicos marcó la pauta en este ámbito al adoptar en 2019 sus Principios sobre Inteligencia Artificial, que fueron actualizados exhaustivamente en 2024 para responder al auge de los modelos generativos avanzados. Estos principios constituyen el primer estándar intergubernamental en la materia y promueven una tecnología innovadora y confiable que respete incondicionalmente los derechos humanos y los valores democráticos.

La base de este marco descansa en cinco principios basados en valores y en cinco recomendaciones de política pública que abordan dimensiones críticas como el crecimiento inclusivo, el desarrollo sostenible, el bienestar social, la robustez técnica y la transparencia operativa. Para mayo de 2023, la base de datos de políticas nacionales de la OCDE ya catalogaba más de mil iniciativas en más de setenta jurisdicciones que seguían explícitamente estas directrices, lo que evidencia el poder de difusión que poseen las normas no vinculantes cuando logran un amplio consenso intergubernamental.

A pesar de estos avances, la principal debilidad del derecho blando radica en la ausencia de mecanismos de ejecución coercitivos. En un entorno altamente competitivo en el que los incentivos económicos de las grandes corporaciones a menudo colisionan con los compromisos éticos declarados,

depender exclusivamente de la autorregulación voluntaria genera brechas sustanciales en materia de seguridad y protección ciudadana. Esta contradicción fundamental precipitó una transición acelerada hacia el derecho duro, caracterizado por normativas vinculantes cuyo incumplimiento acarrea severas sanciones financieras, la prohibición de la comercialización o la imputación de responsabilidades legales.

La interconexión dialéctica entre ambas formas de derecho sugiere que el derecho blando actúa como el precursor necesario del derecho duro. Muchos de los requisitos de gobernanza que hoy son de carácter obligatorio en las legislaciones más avanzadas del mundo proceden directamente de las directrices éticas no vinculantes redactadas años atrás. Esta transición orgánica permite que las leyes definitivas posean un grado de madurez técnica que los legisladores difícilmente habrían alcanzado si hubiesen optado por regular de manera reactiva y prematura en las fases iniciales de la tecnología.

El convenio marco del Consejo de Europa: entre la ambición y la transacción política

Uno de los hitos más significativos en el tránsito del derecho blando al derecho duro a nivel internacional es el Convenio Marco del Consejo de Europa sobre Inteligencia Artificial, Derechos Humanos, Democracia y el Estado de Derecho. Las negociaciones de este histórico instrumento se abrieron formalmente en la primavera de 2022 y culminaron con su apertura para la firma en septiembre de 2024, lo que lo posiciona como el primer tratado internacional jurídicamente vinculante en el ámbito de la gobernanza

algorítmica (Fierro, 2024).

A diferencia de los marcos orientadores precedentes, este convenio impone obligaciones legales estrictas a los Estados que decidan ratificarlo, forzándolos a garantizar de manera efectiva que los sistemas tecnológicos se desarrollen e implementen con el más estricto respeto a los derechos humanos y la preservación de los valores democráticos.

El análisis conceptual de este tratado a través del prisma analítico del proyecto ENSURED —el cual evalúa a las instituciones de gobernanza global bajo las dimensiones de robustez, efectividad y democracia— ofrece una visión matizada de su verdadero potencial y de sus limitaciones intrínsecas. El proceso de negociación, que congregó a gobiernos miembros del Consejo de Europa, estados observadores extracomunitarios, la Unión Europea, representantes del sector privado y organizaciones de la sociedad civil, refleja tanto la ambición de establecer un estándar de alcance global como las inevitables transacciones políticas que debieron realizarse para alcanzar el consenso de los firmantes.

El verdadero impacto del convenio está sujeto a una tensión dialéctica entre su resiliencia institucional y su capacidad real de entrega. Las concesiones realizadas durante las negociaciones derivaron en exenciones explícitas respecto de la regulación de los sistemas de inteligencia artificial operados en el sector privado y de aquellos aplicados directamente a la seguridad nacional.

Para los críticos de la gobernanza blanda, estas exenciones representan una claudicación que debilita significativamente la efectividad del tratado,

dado que gran parte de los riesgos más severos para la privacidad de los datos y el sesgo algorítmico proviene precisamente de las corporaciones tecnológicas privadas y de los sistemas de vigilancia estatal (De Noyette et al., 2025). Por el contrario, para los defensores del pragmatismo regulatorio, estas transacciones políticas fueron necesarias para dotar al tratado de robustez, priorizando su adaptabilidad futura y la accesibilidad para una mayor cantidad de naciones que, de otro modo, se habrían negado a comprometerse bajo reglas demasiado estrictas. La relevancia futura de este instrumento dependerá críticamente de si logra atraer ratificaciones más allá de las fronteras europeas y de la rigurosidad con la que los Estados transpongan sus mandatos a sus ordenamientos jurídicos nacionales.

Convergencia y divergencia en las filosofías de regulación global

La acelerada proliferación de leyes en distintas jurisdicciones ha revelado que los Estados no comparten una visión unívoca sobre cómo debe estructurarse la gobernanza algorítmica. El análisis de las respuestas normativas demuestra que las interpretaciones sobre lo que constituye una regulación de la inteligencia artificial son sumamente amplias, lo que deriva en enfoques que reflejan las prioridades sociopolíticas, las tradiciones jurídicas y los objetivos de soberanía tecnológica de cada bloque económico (Olugbade, 2025).

La Unión Europea ha liderado el desarrollo normativo mediante un modelo precautorio que sitúa los derechos fundamentales y la dignidad

humana en el epicentro de la legislación. El Reglamento de Inteligencia Artificial de la Unión Europea introduce un sistema exhaustivo de clasificación basado en el riesgo, que prohíbe de manera estricta prácticas consideradas intolerables —como los sistemas de calificación social o la manipulación psicológica— e impone pesadas obligaciones de transparencia y control humano para las aplicaciones designadas como de alto riesgo. Este enfoque garantiza un nivel de protección ciudadana sin precedentes, aunque genera inquietud en los sectores empresariales debido al incremento de los costos de cumplimiento y la posible ralentización de la competitividad frente a economías con entornos regulatorios más flexibles (Ruiz, 2023).

En contraposición, la estrategia histórica de los Estados Unidos se ha caracterizado por privilegiar el liderazgo en investigación y desarrollo, recurriendo a un enfoque sectorial y a marcos de gestión de riesgos mayoritariamente voluntarios. La Orden Ejecutiva sobre el Mantenimiento del Liderazgo Estadounidense en Inteligencia Artificial de 2019 cimentó esta postura al priorizar la inversión federal y la reducción de las barreras burocráticas frente a la imposición de obligaciones rígidas.

Si bien este modelo ha demostrado ser altamente efectivo para fomentar ecosistemas de innovación vibrantes y escalar rápidamente los desarrollos comerciales, traslada la carga de la gestión ética y la mitigación de daños a la autorregulación corporativa y a los litigios posteriores ante los tribunales.

Por su parte, la República Popular China integra la gobernanza algorítmica en sus imperativos de seguridad nacional y de estabilidad social. El enfoque chino combina el fuerte fomento de la escalabilidad de la

inteligencia artificial con sistemas estrictos de licencias obligatorias para las aplicaciones incluidas en su Lista Negativa y una férrea moderación estatal de los contenidos generados por los algoritmos.

La Tabla 4 sintetiza y contrasta las características definitorias de estas tres filosofías dominantes, ilustrando cómo el mismo desafío tecnológico da lugar a arquitecturas regulatorias divergentes según el contexto político y social de cada bloque.

Tabla 4: Filosofías de regulación global en el contexto político

Región	**Eje filosófico central**	**Mecanismos de control**	**Fortalezas e implicaciones**
Unión Europea	Enfoque precautorio y derechos fundamentales	Clasificación por niveles de riesgo y auditorías previas al mercado.	Alta protección de los ciudadanos y generación del efecto Bruselas a nivel mundial.
Estados Unidos	Liderazgo de mercado y autorregulación	Marcos de riesgo voluntarios (NIST) y regulaciones sectoriales específicas.	Aceleración de la innovación comercial y una menor carga burocrática inicial.
China	Control estatal y alineación nacional	Moderación estricta de contenidos y licencias mediante listas negativas.	Gran capacidad de escalabilidad industrial bajo estricta supervisión política.

Esta divergencia regulatoria impone una carga formidable a las empresas que aspiran a operar a nivel transfronterizo, obligándolas a tomar decisiones estratégicas sobre si fragmentar sus productos para satisfacer las demandas locales o adoptar el estándar más riguroso disponible como marco operativo universal. El riesgo de una fragmentación tecnológica extrema, impulsada por un nacionalismo algorítmico en el que cada nación impone barreras insalvables para proteger su soberanía digital, representa una de las mayores amenazas para la economía digital hacia 2030.

La auditoría algorítmica y el estándar internacional ISO/IEC 42001

La transición hacia leyes duras y exigibles ha transformado la auditoría de algoritmos en la herramienta operativa primordial para verificar la idoneidad técnica y ética de los sistemas de inteligencia artificial en producción (Oncioiu y Bularca, 2025). Una auditoría algorítmica rigurosa ya no puede limitarse a una simple evaluación estática del código matemático o de la precisión del modelo en entornos de laboratorio controlados. Por el contrario, la práctica profesional exige un enfoque sociotécnico de extremo a extremo que reconozca que los algoritmos operan sobre datos producidos por sociedades humanas complejas y que sus decisiones se despliegan en contextos organizacionales en los que existen profundas asimetrías de poder.

La auditoría algorítmica somete a inspección minuciosa la procedencia y la representatividad de los datos de entrenamiento para evitar que el sistema perpetúe sesgos discriminatorios preexistentes, analiza la lógica interna de

toma de decisiones para garantizar la explicabilidad ante usuarios no técnicos y evalúa el impacto real que produce el despliegue del sistema en las poblaciones afectadas (Baz, 2021). La opacidad intrínseca de los modelos de aprendizaje profundo —comúnmente denominados cajas negras— exige que los auditores implementen métodos avanzados de explicabilidad e interpretabilidad para desentrañar cómo se ponderaron las variables que condujeron a una decisión automatizada con consecuencias significativas para las personas.

Ante la dispersión de criterios de evaluación y la necesidad de contar con metodologías de control uniformes con validez internacional, la Organización Internacional de Normalización publicó en diciembre de 2023 el estándar ISO/IEC 42001.· Esta norma establece los requisitos sistemáticos para que las organizaciones estructuren, mantengan y mejoren de manera continua un Sistema de Gestión de Inteligencia Artificial enfocado en el desarrollo y el uso responsable de la tecnología.

La ISO/IEC 42001 comparte la misma estructura vertebral de alto nivel que otros estándares ampliamente reconocidos por las empresas en materia de gestión de la calidad y de la seguridad de la información, lo que facilita su integración orgánica con los procesos corporativos ya existentes y evita la duplicación innecesaria de estructuras de control (Mazzinghy et al., 2025).

La adopción de este estándar dota a las organizaciones de una infraestructura de gobernanza robusta que no solo mitiga los riesgos legales y reputacionales derivados de fallos algorítmicos, sino que además sirve como un excelente marco de preparación práctica para el cumplimiento de las

severas exigencias que impone el Reglamento de Inteligencia Artificial de la Unión Europea y otras leyes en desarrollo.

La Tabla 5 sintetiza los cinco pilares operativos de control que las organizaciones deben acreditar documentalmente para demostrar su alineación efectiva con los principios de este estándar internacional.

Tabla 5: Pilar de control ISO 42001 en las organizaciones

Pilar de control ISO 42001	**Exigencia operativa en el ciclo de vida**	**Evidencias y registros requeridos**
Gestión de riesgos	Metodologías de aceptación proporcionales al propósito y peligros del sistema	Registros de riesgos vinculados a los documentos de diseño y revisiones de la dirección
Calidad y gobernanza de datos	Evaluación de integridad, representatividad y rastreo de linaje de los datos	Inventarios de fuentes de datos, resultados de calidad y aprobaciones de uso
Transparencia y control humano	Definición de roles de supervisión y diseño de manuales de operación claros	Guías de usuario, playbooks de supervisión y pruebas de anulación del sistema
Robustez y ciberseguridad	Resiliencia proporcional al riesgo y protocolos de manejo de vulnerabilidades	Informes de penetración, reportes de equipos rojos y registros de incidentes
Monitoreo posterior al mercado	Vigilancia activa de los sistemas desplegados y reporte de incidentes	Cuadros de mando de monitoreo y registros de acciones correctivas

	graves	aplicadas

El valor de la certificación bajo la norma ISO/IEC 42001 radica en que permite a las empresas externalizar la confianza de sus clientes e inversores mediante una validación independiente realizada por entidades acreditadas. Esta dinámica promueve un entorno de gobernanza ágil y escalable, indispensable para que las corporaciones naveguen con éxito por las exigencias intrincadas de la transición regulatoria (Mazzinghy et al., 2025).

La brecha estructural en contextos humanitarios y el modelo de cocreación ascendente

Una de las críticas más agudas a la actual arquitectura de gobernanza global radica en su incapacidad para traducirse de manera efectiva en contextos definidos por la extrema vulnerabilidad, la inestabilidad institucional y las profundas asimetrías de poder, siendo el sector de la acción humanitaria el ejemplo más representativo de esta falla estructural.

En estos entornos, los sistemas de inteligencia artificial ya operan como infraestructura crítica activa, determinando quién recibe ayuda alimentaria, quién es catalogado como población de riesgo o quién queda excluido de los recursos de emergencia. A pesar de la gravedad que revisten estas decisiones automatizadas sobre las vidas humanas, no existe un marco de gobernanza sectorial que garantice la seguridad operativa ni la protección de los derechos

de las poblaciones afectadas.

El análisis detallado contenido en el documento inaugural de SAFE AI establece que la brecha de gobernanza en el ámbito humanitario es de naturaleza estructural y no puede resolverse mediante la adopción aislada de políticas institucionales por parte de las agencias de ayuda, por muy rigurosas que estas sean. Los marcos normativos hegemónicos —como el Reglamento de Inteligencia Artificial de la Unión Europea o los Principios de la OCDE— fueron diseñados pensando en los Estados, los reguladores de mercados desarrollados y las corporaciones tecnológicas, por lo que resultan inaplicables en las condiciones severas de los entornos operativos de crisis (Aguirre, 2025).

El sector humanitario enfrenta cuatro desafíos estructurales que los marcos globales no logran abordar de manera satisfactoria: el problema de doble uso de los sistemas, los ecosistemas de información hostiles donde la verificación es casi nula, la asimetría profunda en la rendición de cuentas donde las agencias responden ante los donantes y no ante las poblaciones afectadas, y la responsabilidad sin control operativo, donde las organizaciones humanitarias dependen de infraestructuras de nube concentradas y modelos propietarios que no pueden escrutar ni auditar de manera independiente. Para superar esta crisis de rendición de cuentas, se propone la adopción de infraestructuras de aseguramiento compartido que permitan auditorías independientes de los sistemas antes de su despliegue en misiones críticas.

Frente a la desconexión que exhiben los modelos de gobernanza diseñados desde las altas esferas burocráticas y corporativas, emerge con fuerza la necesidad de adoptar modelos de cocreación ascendente en el diseño

de las políticas y de la propia tecnología. El pensamiento dominante asume erróneamente que los expertos técnicos y los legisladores pueden solucionar los problemas de sesgo y discriminación mediante revisiones matemáticas de los conjuntos de datos, sin comprender que las raíces de la exclusión algorítmica se hunden en las inequidades estructurales de la sociedad.

La participación activa de las comunidades minorizadas en el diseño regulatorio no constituye un acto de concesión ética, sino una exigencia de eficacia operativa. Las personas que pertenecen a comunidades marginadas comprenden las dinámicas de poder y las formas de discriminación que la tecnología puede amplificar mucho mejor que cualquier ingeniero de software o funcionario gubernamental (Oncioiu y Bularca, 2025).

Un ejemplo exitoso de esta metodología de abajo hacia arriba se documentó en Pittsburgh, donde científicos de la computación se aliaron con residentes afectados por la contaminación industrial. Al carecer de datos oficiales para exigir la atención de las autoridades, la comunidad colaboró activamente en el diseño de sistemas de recopilación de información que permitieron presentar reportes ciudadanos sobre olores y datos de calidad del aire ante la Agencia de Protección Ambiental, lo que finalmente forzó el cierre de una fábrica contaminante.

Este enfoque exige que las universidades, los fondos de investigación y las empresas tecnológicas cedan parte de su autoridad a las comunidades y financien su participación voluntaria en los procesos de diseño regulatorio y tecnológico. Organizaciones de la sociedad civil y programas académicos actúan como puentes de traducción entre los tecnólogos y los ciudadanos, impidiendo que el despliegue de tecnologías injustificadas profundice la

vigilancia y la exclusión sobre los sectores más desprotegidos de la sociedad.

El horizonte latinoamericano: Avances normativos en Perú y Brasil

En el contexto geográfico de América Latina y el Caribe, el estado de preparación y gobernanza de la inteligencia artificial exhibe brechas estructurales significativas, caracterizadas por una baja inversión relativa —la región capta apenas el 1,12% de la inversión global en la materia a pesar de representar el 6,6% del Producto Interno Bruto mundial— y una preocupante fuga de talentos especializados hacia mercados desarrollados. El Índice Latinoamericano de Inteligencia Artificial de 2025, elaborado con el análisis de la CEPAL, constata que los países de la región se dividen en tres estadios de madurez: los pioneros —donde Chile, Brasil y Uruguay lideran con puntajes superiores a 60—, los adoptantes y los exploradores.

Si bien el uso de herramientas de inteligencia artificial generativa se ha democratizado rápidamente entre las pequeñas y medianas empresas latinoamericanas debido a su bajo requerimiento técnico inicial, la mayoría de los países carece de estrategias nacionales dotadas de presupuestos adecuados y de mecanismos para evaluar el impacto ético y social (Caiza et al., 2024).

Perú se ha posicionado como uno de los líderes normativos indiscutibles de la región al promulgar el 5 de julio de 2023 la Ley 31814, cuyo objeto explícito es promover el uso de la inteligencia artificial en favor del desarrollo económico y social del país en un entorno seguro que garantice su uso ético y

el respeto irrestricto a los derechos humanos. La ley sienta las bases de la gobernanza peruana sobre principios rectores de seguridad basados en riesgos, pluralidad de participantes, gobernanza de internet y el desarrollo ético de la industria (Carrasco, 2025).

Posteriormente, el 9 de septiembre de 2025 se publicó en el Diario Oficial El Peruano el Decreto Supremo 115-2025-PCM, que aprobó el reglamento formal de la ley, disponiendo su entrada en vigor completa a partir del 22 de enero de 2026 y estableciendo las obligaciones específicas para los actores del Sistema Nacional de Transformación Digital.

El reglamento peruano clasifica las aplicaciones de inteligencia artificial en tres categorías de riesgo e impone severas restricciones a las clasificadas como de uso indebido, por atentar contra la dignidad de las personas y los derechos fundamentales. Para garantizar que la regulación de la inteligencia artificial no asfixie prematuramente el tejido empresarial del país ni frene las capacidades de innovación local, el ordenamiento peruano estableció un calendario de implementación gradual y escalonado.

La Tabla 6 detalla los plazos máximos de cumplimiento obligatorio que deben observar las organizaciones del sector privado en el Perú, de acuerdo con sus áreas específicas de actividad económica.

Tabla 6: Plazos máximos de cumplimiento obligatorio según el sector productivo

Sectores de actividad económica	Límite máximo para la implementación obligatoria
Salud, educación, justicia, seguridad, economía y finanzas	Hasta el 10 de septiembre de 2026
Transporte, comercio y trabajo	Hasta el 10 de septiembre de 2027
Producción, agricultura, energía y minería	Hasta el 10 de septiembre de 2028
Todos los demás sectores y usos no contemplados previamente	Hasta el 10 de septiembre de 2029

Para los sistemas de inteligencia artificial catalogados como de alto riesgo, el marco legal peruano exige que las empresas implementen de manera obligatoria mecanismos de transparencia algorítmica para informar a los usuarios y garantizar la supervisión humana efectiva en la toma de decisiones (Azuaje y Finol, 2020). El personal encargado de esta supervisión debe estar debidamente capacitado para evitar sesgos en la interpretación de los resultados y contar con la potestad de invalidar o anular las decisiones automatizadas emitidas por el algoritmo.

Paralelamente a estos avances reglamentarios, en noviembre de 2025 se presentó ante el Congreso de la República del Perú una propuesta de modificación legislativa impulsada por la congresista Elizabeth Medina, que busca endurecer la fiscalización mediante la creación de un Registro Nacional

de Sistemas de Inteligencia Artificial de Alto Riesgo, con inscripción obligatoria previa a su introducción en el mercado nacional.

En el marco de una convergencia regional, el Senado de la República Federativa de Brasil aprobó el 10 de diciembre de 2024 el Proyecto de Ley 2338 de 2023, que establece las reglas generales para el desarrollo, el fomento y el uso ético de la inteligencia artificial en el gigante sudamericano. El proyecto brasileño sitúa en el núcleo de la gobernanza la centralidad de la persona humana e introduce prohibiciones estrictas sobre las herramientas capaces de inducir conductas peligrosas en poblaciones vulnerables o de realizar la calificación social de los ciudadanos, lo que refleja una clara alineación conceptual con el modelo de protección de derechos humanos propugnado por la Unión Europea.

Tendencias proyectadas y prospectiva de cumplimiento

El horizonte temporal comprendido entre 2026 y 2030 estará definido por tres fuerzas disruptivas convergentes: la explosión sin precedentes de las capacidades algorítmicas, la acelerada madurez de los marcos de cumplimiento y las inevitables fricciones geopolíticas derivadas de la carrera por la supremacía tecnológica. En el plano técnico, la transición hacia el año 2026 marca el advenimiento de modelos que consolidan la multimodalidad como el estándar nativo por defecto y la llegada de agentes autónomos listos para la producción, capaces de ejecutar flujos de trabajo completos sin la mediación humana directa.

Las proyecciones de Gartner anticipan transformaciones profundas en las dinámicas laborales y operativas de las organizaciones en este periodo. Se proyecta que para el año 2027 el 75% de los procesos de contratación laboral incluirán certificaciones o pruebas de alfabetización algorítmica obligatorias para los aspirantes.

El abuso y la sobredependencia de las herramientas generativas generarán una preocupación latente por el deterioro de las capacidades de pensamiento crítico humano, lo que llevará a que, para 2026, la mitad de las organizaciones incorporen de manera sistemática evaluaciones libres de inteligencia artificial en sus procesos de toma de decisiones estratégicas y de selección (Baz, 2021).

Las previsiones apuntan a que se presentarán más de mil demandas legales bajo el concepto de muerte por inteligencia artificial en sectores tan delicados como la medicina asistida por algoritmos, las finanzas y los vehículos autónomos, lo que forzará a las compañías a robustecer sus infraestructuras de gobernanza para evitar quiebras masivas y daños severos a la reputación corporativa.

Se espera, además, que para el año 2028 el 90% de las transacciones de compras entre empresas estén mediadas directamente por agentes algorítmicos autónomos, lo que representará un volumen de intercambio económico superior a los quince billones de dólares y transformará radicalmente las cadenas de suministro globales. Hacia el final de la década, en 2030, se estima que el 22% de las transacciones monetarias mundiales tendrán un carácter programable, dotando a los sistemas de inteligencia artificial de verdadera agencia económica y autonomía financiera.

Esta explosión de capacidades se acompañará de un incremento masivo y sostenido en los costos de cumplimiento normativo del sector empresarial privado. Se estima que el gasto global de las empresas destinado a los departamentos de gobernanza y cumplimiento de inteligencia artificial ascenderá a 2. Mil millones de dólares anuales en 2026, escalando drásticamente hasta alcanzar una proyección de 8. Mil millones de dólares para el año 2034.

Este incremento de costos obligará a las organizaciones a migrar de los análisis de impacto manuales a sistemas de cumplimiento automatizados e impulsados por código, en los que las reglas y restricciones impuestas por las leyes locales se integren de forma nativa como parámetros operativos ineludibles en las tuberías de desarrollo de los ingenieros de software.

El análisis pormenorizado de las dinámicas globales demuestra de manera inequívoca que la gobernanza de la inteligencia artificial ha abandonado para siempre la esfera de las declaraciones de buenas intenciones y de las recomendaciones éticas voluntarias para insertarse firmemente en el campo de la regulación vinculante de derecho duro.

Las implicaciones de este cambio de paradigma resultan formidables para las organizaciones, forzando a los equipos directivos a comprender que el cumplimiento normativo y la gestión profunda de los riesgos no constituyen un freno burocrático para la innovación, sino la única base sólida capaz de otorgar sostenibilidad a largo plazo y de proteger el valor reputacional de las marcas en un mercado altamente vigilado por la sociedad civil y las autoridades estatales (Fernandini y Saavedra, 2025).

La asimetría analizada en los contextos operativos de asistencia humanitaria y la exclusión histórica de las poblaciones más vulnerables en el diseño de las tecnologías dominantes subrayan con urgencia que la legitimidad de las futuras arquitecturas de gobernanza no provendrá únicamente de la sofisticación técnica de los auditores ni de la rigurosidad sancionadora de las leyes vigentes. Por el contrario, la verdadera efectividad de las normas futuras dependerá de la adopción incondicional de modelos de cocreación ascendente, en los que las comunidades afectadas posean una voz vinculante y prioritaria en la definición de qué tecnologías deben construirse y bajo qué salvaguardas operativas inquebrantables se les permitirá operar sobre la vida de los ciudadanos.

El equilibrio que logren alcanzar las naciones entre la urgencia de liderar el desarrollo de las capacidades computacionales y el deber ético e irrenunciable de salvaguardar la dignidad y los derechos fundamentales de las personas marcará indeleblemente el progreso de la civilización humana hacia la mitad del siglo veintiuno.

Capítulo 3

Tensión y convergencia entre la transparencia algorítmica y la propiedad intelectual: marcos regulatorios, técnicos y jurisprudenciales

La integración sistémica de la inteligencia artificial en las estructuras de gobernanza contemporánea ha generado un conflicto fundamental entre dos pilares del ordenamiento jurídico y económico: el derecho a la transparencia y la protección de la propiedad intelectual. En la medida en que los sistemas de toma de decisiones automatizadas (ADM) y los modelos de inteligencia artificial generativa (GenAI) asumen roles críticos en la distribución de recursos, la administración de justicia y la prestación de servicios, la demanda de un escrutinio público de sus lógicas internas se vuelve imperativa para garantizar los derechos humanos y la rendición de cuentas (Baz, 2021).

No obstante, este requerimiento de apertura colisiona directamente con el régimen de propiedad intelectual, en particular con el secreto comercial, que constituye el principal incentivo económico para la innovación tecnológica. Esta tensión no es meramente técnica, sino que representa una disputa sobre la soberanía algorítmica y la capacidad de las democracias para supervisar

poderes tecnológicos opacos (Delva y Mendez, 2025).

Fundamentos de la Tensión: El Velo de la Caja Negra frente al Derecho a Saber

La opacidad algorítmica, a menudo denominada caja negra, se debe a la complejidad inherente a los modelos de aprendizaje profundo (Deep Learning), en los que las interacciones entre millones de parámetros hacen que la trazabilidad de una decisión específica resulte extremadamente difícil, incluso para sus propios desarrolladores. Esta opacidad se ve exacerbada por una opacidad jurídica deliberada, en la que las empresas invocan la propiedad intelectual para restringir el acceso al código fuente, a los pesos del modelo y a los conjuntos de datos de entrenamiento (Morales, 2021).

Desde una perspectiva ética y de derechos humanos, la falta de transparencia genera incertidumbre sobre la equidad y la justicia de los sistemas. Sin explicabilidad, los usuarios afectados no pueden impugnar decisiones discriminatorias o erróneas, lo que debilita la integridad profesional de los desarrolladores y la confianza pública en la tecnología. El derecho a la información, amparado por instrumentos internacionales como la Declaración Universal de Derechos Humanos, exige que cualquier herramienta que tenga un impacto significativo en la vida de las personas sea auditable y comprensible.

En contraste, la teoría económica de la propiedad intelectual sostiene que la divulgación forzada podría desincentivar la inversión, al permitir a competidores realizar ingeniería inversa sin incurrir en los costos de

investigación y desarrollo.

La protección de los algoritmos ha transitado por diversos regímenes jurídicos, adaptándose a la naturaleza intangible y funcional del software. Tradicionalmente, se han utilizado tres vías principales (véase la Tabla 7).

Tabla 7: Evolución de los mecanismos de protección de la propiedad intelectual en IA

Mecanismo de PI	**Objeto de protección**	**Nivel de Transparencia Requerido**	**Desafíos en la IA**
Secreto Comercial	Datos, pesos, arquitecturas, protocolos de entrenamiento.	Nulo; la protección depende de mantener el secreto.	Vulnerable a la ingeniería inversa y a las fugas de datos; en conflicto con las auditorías.
Derechos de autor	Expresión del código fuente y documentación técnica.	Bajo: protege la forma, no la funcionalidad subyacente.	No protege la lógica algorítmica ni los procesos matemáticos.
Patentes	Invenciones técnicas que implementan IA (Core AI).	Alto: requiere una divulgación detallada para habilitar a un experto.	Dificultad para demostrar la novedad y el paso inventivo en modelos evolutivos.

En la práctica actual, el secreto comercial se ha consolidado como el motor silencioso de la industria, ya que ofrece protección inmediata, flexible e

indefinida, sin los costos de registro ni los requisitos de divulgación de las patentes. Esto crea un entorno en el que la innovación se protege mediante el ocultamiento, lo cual choca frontalmente con las nuevas normativas de transparencia.

El régimen de la Unión Europea: el Reglamento de Inteligencia Artificial (AI Act)

La Ley de IA de la UE representa el esfuerzo regulatorio más ambicioso para armonizar la transparencia con la propiedad intelectual. Su enfoque basado en el riesgo establece obligaciones diferenciadas que no buscan asfixiar la innovación, sino proteger los derechos fundamentales (Morales, 2021).

Transparencia Escalonada y Modelos de Propósito General (GPAI)

El Reglamento introduce un marco de transparencia que opera en tres niveles críticos:

1. **Transparencia de cara al usuario (Artículo 50):** Obliga a informar cuando se interactúa con una IA o se visualiza un deepfake. Este nivel tiene un impacto mínimo en los secretos comerciales.
2. **Documentación para sistemas de alto riesgo (Artículo 13):** exige que los proveedores entreguen instrucciones claras y documentación técnica sobre el propósito, la precisión y los datos de entrenamiento a las autoridades y a los usuarios profesionales.

3. **Obligaciones para GPAI (Artículo 53):** Los proveedores de modelos de propósito general deben publicar un resumen suficientemente detallado de los datos utilizados para el entrenamiento. Este punto es el epicentro del conflicto, ya que obliga a levantar el velo sobre la sangre del modelo: su composición de datos.

El artículo 53 busca equilibrar el respeto al derecho de autor con la protección del secreto comercial. Sin embargo, la industria argumenta que incluso un resumen de alto nivel puede revelar ventajas competitivas, como la proporción entre datos licenciados y públicos, la especialización por dominio y los estándares de calidad de los conjuntos de datos propietarios. Para mitigar este riesgo, la Comisión Europea ha propuesto plantillas de documentación que permiten descripciones narrativas en lugar de listados granulares, a fin de preservar la confidencialidad técnica.

El problema estructural del artículo 78: la opacidad defensiva

A pesar de las aspiraciones de transparencia, el artículo 78 del AI Act impone a las autoridades competentes el deber de proteger la información confidencial conforme a las leyes de la Unión. En la práctica, esto crea un desequilibrio: las empresas pueden invocar secretos comerciales con escasa evidencia, mientras que los reguladores deben demostrar que el acceso a datos sensibles es estrictamente necesario para la supervisión. Para Mylly (2023), este estándar no definido por la ley permite a los proveedores retrasar las auditorías mediante reclamaciones de confidencialidad, lo que se ha denominado una estrategia de opacidad defensiva que obstaculiza la ejecución

efectiva del reglamento.

Paisajes Jurisprudenciales: El acceso al algoritmo en los tribunales

Los tribunales están definiendo los límites de la propiedad intelectual cuando esta se utiliza para bloquear el acceso a información de interés público. Dos casos emblemáticos en Europa y uno en Estados Unidos ilustran esta tendencia (Fierro, 2024).

La Doctrina Dun & Bradstreet y el secreto comercial en la UE

El Tribunal de Justicia de la Unión Europea (TJUE), en el asunto *CK v Magistrat der Stadt Wien* (2025), estableció que el secreto comercial no puede constituir un obstáculo categórico al ejercicio del derecho a conocer la lógica de un algoritmo bajo el RGPD. El fallo determina que el responsable del tratamiento debe proporcionar información comprensible sobre los criterios y los datos de entrada, y que solo un tribunal o autoridad independiente —no el propio desarrollador— puede ponderar si el secreto comercial debe ceder ante el derecho de acceso del interesado (Toche, s/f). Esta sentencia es crucial porque retira a las empresas la potestad de ser jueces y parte en la calificación de su propia información como secreta.

El desafío constitucional de xAI en California

En Estados Unidos, la empresa xAI ha demandado al estado de California para anular la Ley de Transparencia de Datos de Entrenamiento

(TDTA). xAI argumenta que la obligación de divulgar resúmenes de datos de entrenamiento constituye una toma (taking) de propiedad privada sin compensación, en violación de la Quinta Enmienda de la Constitución (The Intelligence, s/f). La empresa sostiene que la curación de datos es su secreto comercial más valioso y que su divulgación destruiría su valor económico. Este caso resalta la tensión entre la autoridad regulatoria del Estado, destinada a proteger al consumidor, y los derechos de propiedad constitucionalmente protegidos.

El escenario en América Latina: Innovación y derechos fundamentales

Latinoamérica ha adoptado un papel activo en la regulación de la IA, integrando principios de transparencia en sus legislaciones nacionales para alinearse con los estándares de la OCDE y la UNESCO.

Perú y la Ley 31814: Hacia un registro nacional de IA

Perú promulgó en 2023 la Ley 31814, que promueve el desarrollo de la IA con principios de ética y transparencia. El reglamento de esta ley, publicado en septiembre de 2025, establece una gobernanza encabezada por la Presidencia del Consejo de Ministros (PCM). Una de las propuestas más innovadoras es la creación de un Registro Nacional de Sistemas de Inteligencia Artificial de Alto Riesgo, en el que los proveedores deberán inscribir sus sistemas de forma obligatoria antes de introducirlos en el mercado, lo que garantizará la trazabilidad y la supervisión humana. Para Carrasco (2025), aunque el marco busca fomentar la innovación competitiva en el sector

privado, establece responsabilidades claras para los desarrolladores ante afectaciones a derechos fundamentales.

Chile: Derecho a la explicación y soberanía de datos

Chile tramita actualmente su Proyecto de Ley de IA (Boletín 16821-19), que propone una clasificación en cuatro niveles de riesgo. El proyecto chileno pone especial énfasis en el derecho a la explicación. Los ciudadanos tendrán el derecho explícito a saber si una decisión (como la negativa de un crédito bancario) fue tomada por una máquina y a exigir explicaciones sobre los motivos de dicha decisión. Además, el proyecto incluye disposiciones sobre la sostenibilidad ambiental de la infraestructura de IA y la creación de capacidades de supercomputación (HPC) para reducir la dependencia tecnológica.

Colombia: Ética, justicia y vigilancia

En Colombia, el Ministerio de Ciencias ha liderado un proyecto de ley que regula la IA con un enfoque en la justicia social y la soberanía tecnológica. La propuesta colombiana es particularmente cautelosa con las tecnologías de vigilancia biométrica y la policía predictiva, exigiendo evaluaciones de impacto sobre los derechos humanos y prohibiendo los sistemas que atenten contra la dignidad humana. Colombia sostiene que el secreto propietario no debe impedir la auditabilidad cuando los sistemas afectan el debido proceso en el ámbito judicial o policial.

Mediación técnica: La inteligencia artificial explicable (XAI)

La XAI surge como el puente tecnológico necesario para resolver la dicotomía entre la transparencia y el secreto comercial. Su objetivo es convertir los modelos de caja negra en sistemas interpretables por humanos, sin necesidad de exponer el código fuente completo ni los datos brutos protegidos por PI.

Métodos de explicabilidad y sus implicaciones legales

La implementación de XAI permite una transparencia funcional que cumple con las obligaciones de rendición de cuentas sin desproteger los activos intangibles de las empresas. Los métodos más utilizados incluyen:

1. **LIME (Local Interpretable Model-Agnostic Explanations):** Genera modelos sustitutos simples para explicar decisiones individuales en la vecindad de un dato específico. Es ideal para cumplir con el derecho del usuario final a recibir una explicación sin revelar la arquitectura global del modelo.
2. **SHAP (SHapley Additive exPlanations):** Utiliza la teoría de juegos para cuantificar la contribución de cada característica a una predicción. Se basa en los valores de Shapley, cuya formulación garantiza una distribución equitativa del pago (el resultado de la predicción) entre los jugadores (las características de entrada):

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!}$$

Esta métrica es estándar en auditorías forenses de ciberseguridad debido a su estabilidad y fidelidad matemática.

3. **Explicaciones Contrafácticas:** describen el cambio mínimo necesario en los datos de entrada para que el modelo tome una decisión diferente. Son fundamentales para el debido proceso, ya que orientan al usuario sobre cómo remediar una situación desfavorable (por ejemplo, mejorar su perfil para obtener un crédito).

La integración de estos métodos en el ciclo de vida algorítmico permite pasar de una transparencia reactiva a una transparencia por diseño, en la que la explicabilidad es una característica intrínseca del sistema y no un añadido posterior forzado por la regulación (Aarab et al., 2025).

El rol de la OMPI y el futuro de la propiedad intelectual

La Organización Mundial de la Propiedad Intelectual (OMPI) ha liderado la conversación global sobre la adaptación del sistema de PI a la era de la IA. Sus documentos de cuestiones (Issue Papers) han evolucionado para abordar la relación simbiótica entre tecnología y política, destacando que el sistema de PI no está pasado de moda, sino que se utiliza más que nunca.

Desafíos en Inventiva y Autoría

La OMPI ha identificado que la IA cuestiona la noción tradicional de autoría e inventiva, centrada en el ser humano. Si bien el consenso actual es que la IA es una herramienta y no un sujeto de derecho, surge la pregunta de si los inventos generados de forma autónoma por la IA deberían ser protegibles. La falta de protección para estos productos podría incentivar el ocultamiento de la intervención de la IA, favoreciendo de nuevo el secreto comercial frente a la divulgación de las patentes, lo que ralentizaría el progreso tecnológico global (Ozmen et al., 2023).

Además, la OMPI ha desarrollado herramientas como el Manual de Referencia y la Lista de Verificación para ayudar a las empresas a mitigar los riesgos de PI al usar IA generativa, como la pérdida involuntaria de confidencialidad al introducir secretos comerciales en los prompts de herramientas públicas (Mylly, 2023).

Auditorías de IA: el tercero de confianza y el IAR

La auditoría algorítmica se consolida como el mecanismo de verificación y control continuo necesario para asegurar que los sistemas funcionen a favor de la sociedad y no en su contra.

El Informe de Evaluación de Impacto Algorítmico (IAR)

El IAR es el pilar de la gobernanza algorítmica en sistemas de alto riesgo. Este documento debe incluir una descripción detallada del sistema, la

identificación de los riesgos para los derechos fundamentales y las medidas preventivas implementadas. La auditoría no solo verifica el cumplimiento normativo, sino que también evalúa la equidad de los resultados, la explicabilidad para usuarios no técnicos y el impacto social del sistema.

En el contexto de la administración pública, organismos como la NCA de los Países Bajos han desarrollado marcos de auditoría basados en cinco pilares: gobernanza, modelo y datos, privacidad, calidad de los ITGC y ética. Estas auditorías deben ser realizadas por equipos multidisciplinares que integren competencias técnicas, legales y éticas para obtener una imagen completa del sistema.

La tensión entre la transparencia algorítmica y la propiedad intelectual no debe resolverse anulando uno de los polos, sino mediante un diseño estratégico de gobernanza. La evidencia sugiere que la transparencia absoluta es inviable técnica y económicamente, pero la opacidad total es inaceptable democráticamente (Araya, 2021).

En el futuro, el modelo dominante será el de una transparencia cualificada. Esto implica:

1. **Divulgación confidencial a reguladores:** Acceso total a los componentes sensibles (código, datos, pesos) por parte de autoridades independientes bajo estrictos protocolos de seguridad, similar a los procedimientos de revisión *in camera* en litigios comerciales.
2. **Explicabilidad funcional para los usuarios:** uso obligatorio de técnicas de XAI que proporcionen el porqué de una decisión sin comprometer el cómo técnico que constituye el secreto industrial.

3. **Auditorías sistémicas independientes:** verificación periódica por terceros de confianza que emitan certificados de cumplimiento sin necesidad de publicación abierta de la propiedad intelectual.

En este equilibrio, la propiedad intelectual seguirá siendo el motor de la inversión, mientras que la transparencia funcionará como el sistema de frenado y dirección necesario para que la inteligencia artificial sea una herramienta de progreso alineada con la dignidad humana y el Estado de Derecho. La soberanía tecnológica de las naciones, especialmente en regiones como América Latina, dependerá de su capacidad para implementar estos marcos de manera que fomenten la innovación local sin renunciar al control democrático sobre los algoritmos que estructuran su realidad social y económica (Beliz, 2025).

Capítulo 4

Impacto de la inteligencia artificial en el desarrollo económico y social global y nacional

La inteligencia artificial se ha consolidado como la tecnología de propósito general más disruptiva del siglo veintiuno, con una capacidad de transformación comparable a la de la llegada de la electricidad o de la máquina de vapor. Este fenómeno no representa meramente una evolución en el procesamiento de datos, sino un cambio de paradigma en la arquitectura de la productividad global y de la organización social.

El análisis de las tendencias actuales indica que la inteligencia artificial, especialmente en sus vertientes generativa y agéntica, tiene el potencial de redefinir la relación entre el capital, el trabajo y el Estado, influyendo directamente en el crecimiento del Producto Interno Bruto y en la calidad de vida de las poblaciones. Sin embargo, la trayectoria de este impacto está intrínsecamente ligada a la capacidad de los países para desarrollar infraestructuras robustas, marcos regulatorios éticos y un capital humano capaz de colaborar con sistemas autónomos (Dumouchel, 2023).

Transformación del sector público y servicios gubernamentales

Los gobiernos están adoptando la inteligencia artificial como una herramienta estratégica para mejorar la eficiencia administrativa, la salud fiscal y la prestación de servicios a los ciudadanos. En los mercados emergentes, se estima que la adopción generalizada de la inteligencia artificial en el sector público podría aumentar la productividad de la administración pública hasta un 3%, lo que se traduciría en un incremento del PIB real de hasta un 4% para 2035 (Ospina y Zambrano, 2023).

Caminos de valor en la administración pública

El despliegue de la inteligencia artificial en el sector público se articula en torno a tres pilares fundamentales: la eficiencia gubernamental, la prestación de servicios públicos y el crecimiento económico nacional. En términos de eficiencia fiscal, el uso de analítica avanzada permite optimizar la recaudación de ingresos, reducir el fraude y la duplicación de gastos y fortalecer la sostenibilidad fiscal a largo plazo.

La inteligencia artificial también actúa como un multiplicador de la calidad de los servicios. Por ejemplo, una autoridad tributaria más capaz financia mejor infraestructura, mientras que procesos de concesión de licencias optimizados fomentan el crecimiento de las pequeñas empresas. En el ámbito de la salud, los sistemas de triaje y diagnóstico asistidos por inteligencia artificial permiten compensar la escasez de personal médico y ampliar el alcance de la atención en áreas rurales.

Casos de éxito y digitalización en el servicio ciudadano

Países como Singapur han liderado la implementación de plataformas integradas como Moments of Life, que unifican 15 servicios de distintas agencias, permitiendo a los padres completar trámites de registro de nacimiento y de subsidios en menos de 15 minutos y reduciendo el tiempo administrativo en un 70%. Asimismo, el chatbot OneService de la misma nación gestiona más de 500.000 consultas al año, con una mejora del 50% en la eficiencia de resolución.

En la India, el sistema biométrico nacional Aadhaar utiliza inteligencia artificial para verificar identidades y detectar fraudes, asegurando que las transferencias directas de beneficios lleguen a los destinatarios correctos de manera eficiente. En el Sudeste Asiático, Tailandia proyecta que su mercado de salud digital crecerá un 15% anual gracias a la integración de soluciones de diagnóstico basadas en inteligencia artificial, mientras que en Papúa Nueva Guinea se exploran plataformas de telemedicina para interpretar radiografías en regiones sin especialistas.

El paradigma de la inteligencia artificial en el Perú: Estrategia y regulación

El Estado peruano ha adoptado un enfoque proactivo en la institucionalización de la inteligencia artificial, reconociéndola como una tecnología esencial para el desarrollo nacional. Este compromiso se

materializa en un marco normativo robusto y en una estrategia nacional orientada a posicionar al país como referente regional.

El marco normativo: Ley N° 31814 y su reglamento

La Ley N° 31814, publicada en julio de 2023, establece las bases para promover el uso de la inteligencia artificial en el marco del proceso nacional de transformación digital. La norma prioriza a la persona y el respeto de los derechos humanos, fomentando un desarrollo económico y social ético, sostenible, transparente y responsable.

El marco regulatorio se ha complementado con hitos fundamentales en 2025:

1. **Decreto Supremo N° 115-2025-PCM:** Aprueba el reglamento de la Ley N° 31814, que detalla la aplicación de la norma mediante 36 artículos y seis títulos que guían la adopción tecnológica en los tres niveles de gobierno.
2. **Norma Técnica Peruana sobre Sistemas de Gestión de IA:** Inacal aprobó en julio de 2025 la primera normativa técnica para estandarizar la gestión de la inteligencia artificial en el país.
3. **Secretaría de Gobierno y Transformación Digital (SGTD):** Este órgano actúa como el ente rector encargado de garantizar el uso responsable de la tecnología y de fortalecer la cooperación internacional con líderes globales como Corea del Sur.

La ENIA peruana se fundamenta en seis pilares estratégicos que buscan cerrar las brechas sociales e impulsar la competitividad. Estos ejes están alineados con los objetivos de desarrollo sostenible y consideran la realidad

multicultural del país (véase la Tabla 8).

Tabla 8: La estrategia nacional de inteligencia artificial (ENIA)

Eje estratégico	Objetivo principal	Iniciativas destacadas
Formación	Potenciar el talento humano en todos los niveles.	Programas para 14 000 escolares y docentes en Arequipa.
Modelo Económico	Incorporar la IA en sectores estratégicos y en los servicios públicos.	Uso de IA para detectar tala ilegal (ADETOP v2) por OSINFOR.
Infraestructura	Asegurar datos accionables y capacidad de cómputo.	Construcción del primer Parque Tecnológico Digital en Arequipa.
Datos	Crear infraestructura de datos abiertos y contextualizados.	Bases de datos de lenguas nativas y de la biodiversidad amazónica.
Ética	Garantizar un uso justo, seguro y transparente.	Implementación de los principios de la OCDE y registro público de algoritmos.
Colaboración	Fomentar un ecosistema de innovación internacional.	Centro de Cooperación en Gobierno Digital Corea-Perú.

La visión de Perú hacia 2030 es consolidar el ecosistema nacional de

inteligencia artificial mediante la creación del Centro Nacional de Innovación e Inteligencia Artificial, que actuará como un acelerador de la adopción de la tecnología en las pequeñas y medianas empresas. Actualmente, el país se ubica en la quinta posición en Latinoamérica en el Índice de Preparación para la IA del Gobierno, destacando especialmente por su capacidad de gobernanza y sus políticas públicas.

Aplicaciones sectoriales del Estado peruano

La implementación de la inteligencia artificial en el Perú ya no es teórica, sino que se manifiesta en proyectos concretos que impactan directamente al ciudadano:

- **Salud Pública:** En la Dirección de Redes Integradas de Salud (DIRIS) Lima Este y en hospitales de Lambayeque se utilizan rayos X asistidos por inteligencia artificial para la detección temprana de la tuberculosis y de enfermedades gastroenterológicas. EsSalud La Libertad ha implementado sistemas de triaje de emergencias basados en IA.
- **Justicia:** El Poder Judicial ha lanzado el sistema 'CURIA' y el proyecto piloto 'Mikuna IA', posicionando al sistema jurídico peruano a la vanguardia tecnológica.
- **Medio Ambiente:** El algoritmo ADETOP v2 permite al OSINFOR monitorear la Amazonía desde el espacio para identificar oportunamente actividades de tala ilegal. Asimismo, se desarrollan el primer asistente de biodiversidad y la aplicación ViewLeaf para identificar especies maderables.
- **Seguridad y Educación:** Distritos como San Isidro utilizan IA en sus

centrales de videovigilancia, mientras que el Ministerio de Educación fomenta el talento mediante hackatones de robótica e inteligencia artificial.

Revolución productiva en sectores clave: agricultura y comercio

La inteligencia artificial está redefiniendo la competitividad en sectores tradicionales como la agricultura y el retail, permitiendo una transición de procesos basados en la intuición a decisiones impulsadas por datos masivos en tiempo real.

Inteligencia artificial en la agricultura: hacia la seguridad alimentaria

Se estima que el mercado mundial de la inteligencia artificial en la agricultura crecerá de 2,8 mil millones de dólares en 2025 a 8,5 mil millones en 2030, con una tasa de crecimiento anual compuesta (CAGR) del 25,1%. Esta expansión está impulsada por la necesidad de aumentar la producción de alimentos y optimizar el uso de recursos críticos, como el agua y los fertilizantes.

La agricultura de precisión lidera el mercado con una participación superior al 43%, lo que permite a los agricultores realizar prescripciones automáticas de siembra que reducen el gasto en semillas entre un 8% y un 12%. Las tecnologías de visión computacional, con un crecimiento proyectado anual del 22,68%, están permitiendo el despliegue de flotas de drones y robots

autónomos para la cosecha y el control de malezas (véase la Tabla 9).

Tabla 9: Mercado mundial de la inteligencia artificial en la agricultura

Aplicación agrícola	**Beneficio principal**	**Estado de adopción / Proyección**
Agricultura de precisión	Optimización de insumos y rendimiento	43 % del mercado en 2025.
Monitoreo ganadero	Salud animal y gestión de recursos	Crecimiento sostenido en sensores IoT.
Invernaderos inteligentes	Control ambiental automatizado	CAGR proyectada de 22% hasta 2031.
Análisis de drones	Detección temprana de plagas y estrés hídrico	Expansión de los servicios de imágenes de alta resolución.

A pesar de su potencial, el sector enfrenta desafíos estructurales. La brecha digital es pronunciada: mientras las multinacionales y grandes empresas adoptan rápidamente herramientas como los copilotos de IA generativa para la agronomía, la adopción entre los pequeños agricultores se mantiene por debajo del 5% debido a los altos costos iniciales de los sensores y a la falta de estándares de datos agronómicos. Para que la inteligencia artificial favorezca el desarrollo social en el campo, es imperativo democratizar el acceso mediante subsidios gubernamentales y plataformas en la nube asequibles.

El nuevo paradigma del comercio minorista y la industria

En el comercio minorista, la inteligencia artificial ha pasado de ser un experimento a convertirse en el motor de la rentabilidad. Las empresas que han adoptado sistemas de aprendizaje automático han reportado un crecimiento anual de los beneficios del 8% en 2023 y 2024, superando significativamente a sus competidores no tecnológicos.

La IA agéntica representa el avance más reciente, con sistemas que actúan de forma autónoma para alcanzar objetivos de negocio. Walmart ha implementado agentes de IA para automatizar las negociaciones con proveedores, con una tasa de éxito del 68%, lo que ha permitido lograr ahorros promedio del 3% en los costos de adquisición. En términos de experiencia del cliente, el uso de la personalización impulsada por IA ha incrementado el valor promedio de los pedidos en un 25% y ha reducido las tasas de devolución en un 19%.

La manufactura también experimenta una profunda transformación, con un mercado de IA que se prevé alcanzará los 20,8 mil millones de dólares para 2028. El mantenimiento predictivo es la aplicación estrella: utiliza datos de sensores para prever fallas en la maquinaria, lo que reduce drásticamente el tiempo de inactividad y los costos de reparación.

Inteligencia artificial y los Objetivos de Desarrollo Sostenible (ODS)

La inteligencia artificial tiene una relación simbiótica con los ODS de las

Naciones Unidas. Su capacidad para detectar patrones en grandes conjuntos de datos la hace invaluable para monitorear el medio ambiente y diseñar intervenciones sociales precisas.

Contribución a los pilares sociales: pobreza, salud y educación

Para el ODS 1 (Fin de la Pobreza), la inteligencia artificial puede actuar como un habilitador para el 100% de sus metas. El análisis de imágenes satelitales permite identificar regiones con necesidades críticas de infraestructura, mientras que la analítica financiera facilita el acceso al crédito para los 1,7 mil millones de adultos que aún carecen de servicios bancarios.

En el ámbito del ODS 2 (Hambre Cero), proyectos como NASA Acres y ZeroHungerAI utilizan años de datos satelitales para predecir crisis alimentarias y optimizar las prácticas de cultivo. En el sector educativo (ODS 4), las plataformas de aprendizaje adaptativo están ayudando a los docentes en entornos de bajos recursos a personalizar la enseñanza, permitiendo que cada estudiante aprenda a su propio ritmo (véase la Tabla 10).

Tabla 10: ODS y los pilares sociales

ODS Relacionado	**Aplicación de la inteligencia artificial**	**Impacto observado / potencial**
ODS 3: Salud	Diagnóstico por imagen y descubrimiento de fármacos.	Reducción de la carga administrativa y mayor acceso en zonas rurales.

ODS 7: Energía	Redes inteligentes y optimización de renovables.	Incremento del 20% en la producción eólica mediante monitoreo con software.
ODS 13: Acción Climática	Predicción de desastres y monitoreo de emisiones.	Alertas tempranas para inundaciones e incendios forestales.
ODS 15: Vida de ecosistemas	Monitoreo de la biodiversidad y detección de deforestación.	Identificación de la tala ilegal mediante modelos acústicos de deep learning.

Sostenibilidad ambiental y economía circular

La inteligencia artificial es una pieza clave para la transición hacia una economía circular, que busca diseñar productos sin residuos y mantener los materiales en uso. La tecnología de diseño inverso permite a los investigadores especificar las propiedades deseadas de un material y dejar que la IA proponga la estructura molecular óptima, lo que acelera la generación de alternativas biodegradables.

En la gestión de residuos sólidos, los sistemas de clasificación robótica guiados por visión computacional pueden separar plásticos, metales y vidrios en cintas transportadoras de alta velocidad con una precisión inalcanzable para los humanos, lo que genera corrientes de reciclaje mucho más limpias. Sin embargo, este potencial ecológico se enfrenta a la realidad de la huella ambiental de la propia tecnología. El consumo eléctrico de los centros de datos podría triplicarse para 2030 y la demanda de agua para la refrigeración ya es

motivo de preocupación en regiones con escasez hídrica.

Desafíos éticos, sociales y el futuro del trabajo

A pesar de los beneficios económicos, la inteligencia artificial presenta riesgos significativos de disrupción social, particularmente en el mercado laboral y en la equidad de acceso a los beneficios tecnológicos.

La brecha de IA y la desigualdad laboral

La inteligencia artificial amenaza con profundizar la brecha entre trabajadores calificados y no calificados. Aquellos con habilidades centradas en los datos y la tecnología verán su productividad multiplicada, mientras que los trabajadores en roles rutinarios enfrentan el riesgo de ser desplazados. En los países de altos ingresos, hasta el 5,1% de la fuerza laboral está en riesgo de ser automatizada, frente al 0,4% en los de bajos ingresos, lo que, paradójicamente, refleja una mayor vulnerabilidad en las economías más desarrolladas.

La gestión algorítmica también plantea preocupaciones éticas. En el mercado del trabajo digital (gig economy), los algoritmos a menudo controlan a los trabajadores mediante reglas estrictas, creando sistemas en los que el trabajador es tratado como un recurso reemplazable, con poca transparencia sobre cómo se evalúa su desempeño o cómo se le asignan las tareas.

Los sistemas de inteligencia artificial pueden perpetuar sesgos históricos si se entrenan con datos que excluyen o tergiversan la representación de grupos marginados. Para los pueblos indígenas, esto representa un riesgo de apropiación cultural y de exclusión de los beneficios

del desarrollo digital. La UNESCO y la OCDE han enfatizado la necesidad de una gobernanza inclusiva en la que las comunidades vulnerables no sean solo beneficiarias pasivas, sino también codiseñadoras de las herramientas tecnológicas que las afectan (véase la Tabla 11).

Tabla 11: Sesgos, privacidad y derechos de las comunidades vulnerables

Desafío ético	**Riesgo identificado**	**Recomendación de política**
Sesgo algorítmico	Discriminación en el acceso al crédito o al empleo.	Auditorías de algoritmos y uso de conjuntos de datos representativos.
Privacidad	Vigilancia masiva y uso indebido de datos de salud.	Marcos de protección de datos personales y soberanía digital.
Desplazamiento laboral	Pérdida de empleos en el sector administrativo.	Programas de formación continua y de protección social adaptativa.
Brecha de acceso	Concentración de la capacidad de cómputo en el Norte Global.	Cooperación internacional para el acceso a la infraestructura de IA.

Estrategias para una inteligencia artificial inclusiva y responsable

Para asegurar que la inteligencia artificial actúe a favor del desarrollo económico y social, es fundamental que los responsables de las políticas públicas sigan recomendaciones basadas en la ética y la inclusión. Las mejores prácticas internacionales sugieren un enfoque centrado en el ser humano que priorice la transparencia y la rendición de cuentas (Ozmen et al., 2023).

Recomendaciones de la OCDE y la UNESCO

Los principios de la OCDE sobre la inteligencia artificial, actualizados en 2024, establecen que los actores de la IA deben comprometerse con la transparencia y la divulgación responsable, proporcionando información que permita a las personas comprender los resultados generados por los sistemas inteligentes (Gunasekara et al., 2025). La UNESCO, a través de su Recomendación sobre la Ética de la IA, subraya que la tecnología debe ser una herramienta para cerrar las brechas de género y promover la diversidad cultural.

El programa RAM (Readiness Assessment Methodology) de la UNESCO ha sido implementado en más de 75 países para ayudar a los gobiernos a identificar sus brechas en materia de políticas públicas y a desarrollar estrategias nacionales que aseguren un despliegue ético de la tecnología. Asimismo, el G20 ha impulsado la creación de facilidades de asistencia técnica para ayudar a los países en desarrollo a diseñar marcos regulatorios que eviten la concentración del mercado y fomenten la innovación local.

El rol de la infraestructura y la soberanía digital

El éxito de la inteligencia artificial en el Sur Global depende críticamente del acceso a la infraestructura de cómputo. Actualmente, África posee menos del 1% de la capacidad de centros de datos del mundo, a pesar de albergar al 18% de la población mundial. Sin capacidad local para procesar datos, los países en desarrollo se ven obligados a exportar sus datos y a depender de modelos diseñados fuera de su contexto cultural y lingüístico.

La soberanía de datos emerge como una prioridad estratégica. Los países deben desarrollar la capacidad de entrenar sus propios modelos con datos locales, lo cual es esencial para aplicaciones en salud, justicia y la preservación de lenguas nativas. Esto requiere una inversión masiva no solo en hardware, sino también en redes eléctricas resilientes, ya que el entrenamiento de modelos de frontera consume miles de megavatios-hora.

El análisis de la trayectoria actual de la inteligencia artificial revela una tecnología con una capacidad dual: puede ser el motor más potente para el crecimiento y la equidad social en la historia reciente, o convertirse en el mayor generador de desigualdades si se deja a la inercia del mercado.

La experiencia de países como Perú demuestra que el liderazgo estatal, mediante leyes modernas y estrategias participativas, es vital para canalizar el poder de la inteligencia artificial al servicio del bien público. Los casos de uso en salud, protección forestal y justicia en el país son evidencia de que es posible adaptar tecnologías de vanguardia a las necesidades de poblaciones diversas.

Sin embargo, para que el desarrollo económico sea sostenible y

socialmente justo, se deben abordar con urgencia las brechas de infraestructura y de talento. La inteligencia artificial no sustituirá la necesidad de un gobierno humano sólido, sino que expondrá las deficiencias de quienes carecen de una estrategia clara para adoptarla. La transición hacia una sociedad impulsada por la inteligencia artificial requiere un contrato social renovado que garantice que el aumento de la productividad se traduzca en una mejora tangible de los ingresos de los hogares y en la protección de los derechos de los más vulnerables.

En última instancia, el éxito de la inteligencia artificial en favor del desarrollo no se medirá por la velocidad de sus procesadores o el tamaño de sus modelos, sino por su capacidad para reducir la pobreza, mejorar la salud global y permitir una convivencia armoniosa con los ecosistemas del planeta. La colaboración internacional, la transparencia algorítmica y la educación continua son los únicos pilares que permitirán que esta revolución tecnológica cumpla su promesa de progreso universal (Araya, 2021).

Capítulo 5
El marco de la UNESCO para la gobernanza de plataformas digitales: un análisis sistémico hacia un internet de confianza

La transformación radical del ecosistema global de información en las últimas décadas ha desplazado el centro de gravedad del discurso público hacia infraestructuras privadas transnacionales, conocidas como plataformas digitales. Si bien estas herramientas han democratizado el acceso al conocimiento y facilitado la conexión global, también se han convertido en vectores de riesgos sistémicos que amenazan los derechos humanos y la estabilidad democrática.

Ante esta dicotomía, la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) ha consolidado su papel como la agencia líder en la protección de la libertad de expresión y el acceso a la información mediante la publicación de las Directrices para la gobernanza de las plataformas digitales: Salvaguardar la libertad de expresión y el acceso a la información con un enfoque de múltiples partes interesadas. Este documento, presentado en noviembre de 2023 y actualizado continuamente para reflejar los avances tecnológicos, representa no solo un manual técnico, sino un consenso político y social forjado a través de uno de los procesos de consulta más extensos en la historia de las Naciones Unidas, integrando más

de 10.000 comentarios procedentes de 134 países.

La necesidad de un marco global coordinado se sustenta en datos alarmantes que subrayan la fragilidad del entorno digital actual. Según el Informe de Riesgos Globales del Foro Económico Mundial de 2025, la desinformación y la información errónea se han identificado como los principales riesgos globales a corto plazo por segundo año consecutivo.

Esta percepción se ve respaldada por el hecho de que el Índice Global de Libertad de Expresión ha caído un 10% desde 2012, mientras que el control gubernamental y de grupos de poder sobre los medios tradicionales y digitales ha aumentado un 48% en el mismo periodo. En este contexto, aproximadamente 1. millones de personas ganaron acceso a redes sociales y plataformas de mensajería entre 2022 y 2025, lo que agrava la urgencia de establecer mecanismos de gobernanza que protejan la integridad de la información sin recurrir a la censura arbitraria.

Evolución histórica y fundamentos normativos de la gobernanza digital

La génesis de las directrices actuales se remonta a la Declaración Windhoek+30 de 2021, en la que los Estados miembros de la UNESCO reconocieron que la información es un bien público y establecieron tres pilares fundamentales para la acción: garantizar la transparencia de las plataformas, asegurar la viabilidad de los medios de comunicación y fomentar la alfabetización mediática e informacional (AMI) entre los ciudadanos. El marco de la UNESCO se aparta deliberadamente de los intentos previos de regulación

unilateral o meramente estatutaria, proponiendo en su lugar un enfoque de gobernanza multiactor que involucra a Estados, empresas tecnológicas, organizaciones de la sociedad civil, la academia y la comunidad técnica.

Un aspecto crítico de esta evolución es el cambio de enfoque desde la regulación directa de contenidos hacia la regulación de sistemas y procesos. Históricamente, los intentos de los gobiernos por controlar lo que se publica en internet han derivado con frecuencia en censura o en medidas desproporcionadas, como los apagones de internet (internet shutdowns), que alcanzaron niveles récord en 2024. La UNESCO argumenta que la gobernanza debe centrarse en los mecanismos de diseño de las plataformas —incluidos los algoritmos de recomendación, los sistemas publicitarios y los procesos de moderación— para mitigar los riesgos estructurales antes de que estos se traduzcan en daños individuales o sociales.

La estructura de la gobernanza: coexistencia de mecanismos

El marco de la UNESCO no impone un modelo único, sino que describe un ecosistema en el que deben coexistir diversos arreglos regulatorios, siempre bajo el estándar del derecho internacional de los derechos humanos (véase la Tabla 12). Esta flexibilidad es esencial para evitar la fragmentación de internet, una preocupación creciente en la que las regulaciones nacionales divergentes terminan por erosionar el carácter global y abierto de la red (Sklavos et al., 2024).

Tabla 12: Mecanismos de gobernanza de la UNESCO

Mecanismo de Gobernanza	Descripción y Alcance	Rol del Estado
Autorregulación	Organismos de la industria o consejos de supervisión que gestionan sus propias normas de conducta y sus términos de servicio.	Observador y promotor de mejores prácticas sin intervención directa.
Corregulación	Desarrollo de un conjunto de códigos de conducta entre plataformas, reguladores y la sociedad civil, que pueden contar con respaldo legal.	Facilitador del diálogo y garante de que los acuerdos se alineen con los derechos humanos.
Regulación Estatutaria	Establecimiento de leyes que facultan a reguladores independientes para supervisar procesos y aplicar sanciones.	Creador del marco legal y garante de la independencia de la autoridad reguladora.

Principios fundamentales para plataformas digitales

El núcleo de las directrices establece cinco principios que las plataformas digitales deben cumplir para garantizar un entorno respetuoso de los derechos humanos. Estos principios actúan como una brújula ética y operativa para las empresas, independientemente de su tamaño o modelo de negocio.

El deber de diligencia debida en derechos humanos

Las plataformas tienen la obligación de realizar evaluaciones periódicas de riesgo para identificar y abordar cualquier impacto potencial o real en los derechos humanos derivado de sus operaciones. Estas evaluaciones deben ser especialmente rigurosas antes de cambios significativos en el diseño, de decisiones políticas importantes (como la alteración de sistemas publicitarios) o del lanzamiento de nuevos servicios en regiones con contextos políticos volátiles. Un hallazgo relevante es que las plataformas suelen operar bajo una lógica de control de daños ex post facto, que la UNESCO busca revertir mediante una cultura de prevención sistémica.

Adhesión a estándares internacionales en diseño y moderación

El diseño de la interfaz de usuario, los algoritmos de curación de contenidos y los sistemas de moderación deben alinearse con el derecho internacional, en particular con el Artículo 19 del Pacto Internacional de

Derechos Civiles y Políticos (PIDCP). Esto implica que las plataformas deben invertir en moderación multilingüe para evitar que los usuarios de lenguas menos representadas o indígenas queden desprotegidos frente a discursos de odio o de desinformación. Se estima que una de las mayores fallas actuales de las plataformas es la asimetría lingüística: mientras que el contenido en inglés recibe una supervisión intensiva, el de otros idiomas carece de contexto y de recursos adecuados.

La transparencia no debe limitarse a la publicación de extensos documentos legales que los usuarios rara vez leen. La UNESCO exige transparencia en los procesos algorítmicos, en el funcionamiento de la publicidad dirigida y en la financiación de los contenidos. Además, las plataformas deben proporcionar a los investigadores externos y a la sociedad civil acceso a datos no personales (vetted researchers) para realizar auditorías independientes del impacto social de los algoritmos (véase la Tabla 13).

Tabla 13: Transparencia y acceso a la información

Dimensión de la Transparencia	**Requisito para las plataformas**	**Impacto en el usuario**
Algoritmos	Explicar cómo se prioriza, se recomienda y se amplifica el contenido.	Reducción de la opacidad en las cámaras de eco.

Publicidad	Identificar claramente los anuncios, quién los paga y los criterios de microsegmentación.	Mayor capacidad para detectar manipulaciones políticas o comerciales.
Moderación	Notificar al usuario por qué se eliminó su contenido y qué política específica se infringió.	Garantía del debido proceso y del derecho a la réplica.

Empoderamiento del usuario y herramientas críticas

El marco pone un énfasis sustancial en reducir la brecha de participación mediante la promoción de la alfabetización mediática e informacional. Las plataformas deben estar diseñadas de manera que incorporen principios de AMI y proporcionen herramientas para que los usuarios gestionen su propia experiencia, controlen la recopilación de sus datos personales y comprendan cuándo interactúan con contenido generado por inteligencia artificial.

Rendición de cuentas ante las partes interesadas

Las plataformas deben establecer mecanismos de queja y apelación efectivos, culturalmente sensibles y fáciles de usar. La rendición de cuentas también incluye la obligación de someterse a auditorías externas e independientes que evalúen la conformidad de la plataforma con sus propios

términos de servicio y con los estándares internacionales de derechos humanos (Fernandini y Saavedra, 2025).

El rol del regulador independiente y los sistemas de pesos y contrapesos

Para que la gobernanza digital sea legítima, la UNESCO subraya la necesidad de que las autoridades regulatorias operen con total independencia del poder político y de los intereses comerciales. Un regulador independiente debe contar con autonomía financiera, procesos de nombramiento transparentes y una protección legal estricta contra despidos arbitrarios (Sklavos et al., 2024).

Las facultades de estos organismos deben incluir la capacidad de solicitar información detallada sobre la moderación de contenidos y el uso de datos (especialmente en periodos electorales), elaborar recomendaciones de buenas prácticas y coordinar acciones con otros organismos públicos para proteger los derechos digitales de los ciudadanos. Sin embargo, la UNESCO advierte que el regulador no debe tener el poder de dictar lo que es verdad, sino de supervisar que los procesos de las plataformas para gestionar la información sean transparentes y respetuosos de los derechos.

Cooperación regional y global: El foro global de Redes

Reconociendo que las plataformas operan más allá de las fronteras nacionales, la UNESCO facilitó en 2024 la creación del Foro Global de Redes de Reguladores. Este espacio permite que autoridades de diferentes regiones,

como Europa y América Latina, intercambien experiencias sobre desafíos comunes como la propagación de desinformación electoral y el acoso en línea contra periodistas y mujeres. El foro es crucial para abordar la relación asimétrica entre los Estados y las empresas tecnológicas globales, lo que permite a los países con marcos regulatorios menos desarrollados aprender de las experiencias de jurisdicciones como la Unión Europea (Aguirre, 2025).

Inteligencia artificial y la nueva frontera de la gobernanza

El auge de la inteligencia artificial (IA), en particular la IA generativa, ha introducido una capa adicional de complejidad que la UNESCO aborda mediante un documento complementario a las directrices de gobernanza. La organización sostiene que la IA no debe ser un campo de autorregulación exclusiva, dado su potencial para amplificar sesgos discriminatorios y facilitar la creación de desinformación a gran escala mediante deepfakes.

En 2021, la UNESCO adoptó por unanimidad la Recomendación sobre la Ética de la Inteligencia Artificial, el primer estándar global en la materia. Este documento establece que la gobernanza de la IA debe abarcar todo el ciclo de vida del sistema, desde el diseño hasta el despliegue y la auditoría (véase la Tabla 14). Para la gobernanza de plataformas, esto implica que los algoritmos de aprendizaje automático deben ser auditables, trazables y estar sujetos a una supervisión humana significativa (Delva y Mendez, 2025).

Tabla 14: El nexo entre la recomendación sobre la ética de la IA y la gobernanza de plataformas

Principio de IA ética	Aplicación en plataformas digitales	Riesgo Mitigado
Proporcionalidad	El uso de la IA para la moderación no debe exceder lo necesario para un fin legítimo.	Censura excesiva por falsos positivos algorítmicos.
Transparencia	Derecho del usuario a recibir una explicación de las decisiones automatizadas.	Discriminación opaca en la distribución de contenidos.
Sustentabilidad	Evaluación del impacto ambiental del entrenamiento de modelos de IA.	Contribución desproporcionada al cambio climático por la infraestructura de datos.
Oversight Humano	Existencia de personal capacitado para anular las decisiones de la IA.	Pérdida de responsabilidad legal y ética.

La aplicación del principio de precaución

Ante la incertidumbre sobre las capacidades futuras de la IA generativa, la UNESCO aboga por el principio de precaución. Esto implica que las empresas deben priorizar las medidas preventivas y mitigar los riesgos antes de que los daños se materialicen, especialmente en lo que respecta a la desinformación electoral y a la protección de la privacidad infantil. Un ejemplo de implementación práctica es el uso de sandboxes regulatorios, entornos controlados en los que los desarrolladores de IA pueden probar sus sistemas bajo supervisión gubernamental antes de su lanzamiento comercial masivo.

Implementación regional: América Latina y África como laboratorios de cambio

La UNESCO ha centrado sus esfuerzos de implementación en el Sur Global para reducir la brecha en la capacidad regulatoria. A través de proyectos financiados por la Comisión Europea, se están fortaleciendo las competencias de las autoridades y de la sociedad civil en países del Cono Sur y de África.

El liderazgo de Colombia en IA judicial

Colombia se ha posicionado como un referente global al ser el primer país en adaptar las directrices de la UNESCO sobre el uso de la IA en el sistema judicial. Tras una sentencia de la Corte Constitucional (T-323/24) que alertó sobre el uso de ChatGPT en fallos judiciales sin regulación, el Consejo Superior de la Judicatura lanzó en diciembre de 2024 un marco que diferencia

claramente entre aplicaciones de bajo y de alto riesgo. El modelo colombiano exige que los operadores judiciales revelen siempre el uso de herramientas de IA en sus decisiones y sostiene que la razón humana es insustituible.

El modelo brasileño de lucha contra la desinformación

Brasil ha implementado respuestas institucionales innovadoras, como el Centro Integrado de Lucha contra la Desinformación y Defensa de la Democracia (CIEDDE) del Tribunal Superior Electoral (TSE). Este centro promueve la cooperación entre el poder judicial, los organismos públicos y las plataformas digitales para garantizar el cumplimiento de las normas electorales y combatir la monetización de canales que difunden noticias falsas sobre el sistema de votación. Asimismo, la Autoridad Nacional de Protección de Datos (ANPD) de Brasil ha lanzado un Sandbox Regulatorio para IA y Protección de Datos que funcionará hasta 2026 y permitirá el desarrollo de innovaciones éticas.

Desafíos en el continente africano

En África, la UNESCO está impulsando la gobernanza de datos mediante la iniciativa ICEGOV2025 y colaboraciones con la Unión Africana. El desafío principal en esta región radica en pasar del cumplimiento regulatorio básico a un empoderamiento real que refleje la diversidad cultural y lingüística del continente. Se subraya la necesidad de institucionalizar prácticas de gobernanza que no solo regulen, sino que también fomenten la infraestructura técnica y el talento local para evitar una nueva forma de dependencia tecnológica.

Críticas y perspectivas de la sociedad civil: Un diálogo inacabado

A pesar del reconocimiento global del marco de la UNESCO, diversas organizaciones de la sociedad civil han expresado preocupaciones legítimas sobre los riesgos de un uso indebido de las directrices. Entidades como Global Partners Digital (GPD), Access Now y la Electronic Frontier Foundation (EFF) han señalado que, en contextos no democráticos o con sistemas de justicia frágiles, el lenguaje sobre la regulación estatal de plataformas podría interpretarse como una luz verde para imponer restricciones desproporcionadas a la libertad de expresión.

Riesgos de abuso y censura estatal

Una de las críticas centrales es la falta de claridad inicial sobre qué constituye exactamente un daño significativo a la democracia, lo que podría permitir que los gobiernos autoritarios etiqueten como desinformación cualquier discurso crítico o de oposición. En respuesta a estas críticas, la versión final de las directrices de 2023 reforzó la necesidad de una supervisión judicial independiente y especificó que la gobernanza no debe debilitar las tecnologías de protección de la privacidad, como el cifrado de extremo a extremo.

La asimetría Norte-Sur en la moderación de contenidos

Desde una perspectiva de economía política, se ha cuestionado que el marco no aborde adecuadamente las desigualdades en la industria de la

moderación. Los moderadores de contenido, con frecuencia ubicados en el Sur Global, realizan el trabajo traumático de limpiar las redes para los usuarios del Norte Global, mientras que los contenidos que circulan en sus propios idiomas nacionales a menudo carecen de la misma protección algorítmica y humana. La UNESCO ha incorporado este desafío, exigiendo que las plataformas demuestren su efectividad en la mitigación de riesgos en todos los idiomas y contextos en los que operan, aunque la aplicación práctica de esta demanda sigue siendo un reto.

Integración en la agenda multilateral global

El marco de la UNESCO es una pieza fundamental de un rompecabezas más amplio: la arquitectura de la cooperación digital de las Naciones Unidas. Existe una alineación estratégica entre estas directrices y el Global Digital Compact (GDC), adoptado en septiembre de 2024 como parte del Pacto para el Futuro.

El Global Digital Compact constituye el primer instrumento multilateral integral para la gobernanza de la IA y la cooperación digital. Mientras que el GDC establece los principios políticos de alto nivel —como cerrar la brecha digital y promover un internet abierto y seguro—, las directrices de la UNESCO proporcionan las herramientas operativas y las normas específicas para que estos objetivos se traduzcan en políticas públicas nacionales y en prácticas corporativas (Cihon et al., 2021). De igual manera, las directrices ecoan el Código de Conducta de la ONU para la Integridad de la Información en Plataformas Digitales, buscando una respuesta coherente de todo el sistema de las Naciones Unidas frente a las amenazas informativas globales.

El marco genérico de la UNESCO para la gobernanza digital marca el fin de la era de la autorregulación absoluta y el inicio de una fase de responsabilidad compartida. La premisa fundamental es que el mercado, por sí solo, no puede garantizar la integridad de la información ni proteger los derechos humanos ante modelos de negocio que monetizan la atención mediante la polarización (Sklavos et al., 2024).

Para que este marco sea efectivo a largo plazo, se requieren acciones sostenidas en tres frentes:

1. **Fortalecimiento Institucional:** Los Estados deben garantizar la independencia técnica y financiera de sus reguladores digitales para evitar la captura política.
2. **Inversión en Alfabetización:** La gobernanza no es solo una cuestión de leyes, sino también de ciudadanos capaces de discernir y participar críticamente en el entorno digital.
3. **Vigilancia de la Sociedad Civil:** Las organizaciones no gubernamentales deben actuar como perros guardianes (watchdogs), monitoreando tanto el comportamiento de las plataformas como el de los gobiernos para asegurar que la regulación no se convierta en censura.

La UNESCO ha concebido sus directrices como un documento vivo que debe revisarse periódicamente para adaptarse a innovaciones como la computación cuántica o las nuevas interfaces cerebro-computadora. El éxito final del marco no se medirá por la cantidad de regulaciones creadas, sino por la recuperación de la confianza de los ciudadanos en los espacios digitales, asegurando que internet siga siendo un motor para el desarrollo sostenible,

la paz y la dignidad humana.

Conclusión

La investigación concluye que no hay un conflicto absoluto, sino una tensión dinámica. Mientras que el secreto empresarial impulsa la innovación y protege la ventaja competitiva, la opacidad de los algoritmos (el fenómeno de la caja negra) puede poner en riesgo derechos fundamentales como la igualdad, el debido proceso y la no discriminación.

Los marcos existentes de propiedad intelectual e industrial, en particular las leyes sobre secretos comerciales, no estaban diseñados para abordar la complejidad de la inteligencia artificial. La protección automática y de duración indefinida del secreto empresarial entra en conflicto con la necesidad de auditorías públicas, especialmente cuando los algoritmos se emplean en ámbitos sensibles como la justicia, la salud o la contratación laboral.

Se ha identificado que la transparencia no requiere necesariamente la revelación del código fuente. Existen diversos niveles:

- Transparencia técnica: Acceso al código (reservado para expertos/auditores).
- Transparencia explicativa: información sobre la lógica, los datos de entrenamiento y los criterios de ponderación (dirigida al usuario final).
- Transparencia procedimental: información sobre quién diseñó el sistema y para qué fue diseñado.

La protección de los activos inmateriales de una empresa debe limitarse cuando hay un riesgo comprobado para el interés público o los derechos humanos. La jurisprudencia reciente indica que el derecho a la explicabilidad

es un requisito previo para ejercer otros derechos en una sociedad democrática.

Con base en estos hallazgos, se recomienda:

- *Para el marco regulatorio (legisladores):*

- Implementar la transparencia por diseño: exigir que los sistemas de IA de alto riesgo incluyan mecanismos de registro (logs) y de trazabilidad que permitan auditorías sin exponer el núcleo del secreto comercial.
- Definir excepciones de interés público: legislar cláusulas claras que faculten a las autoridades de supervisión para acceder a la lógica algorítmica bajo estrictos protocolos de confidencialidad cuando se sospechen sesgos o daños.
- Armonización con el reglamento de IA: alinear las medidas de protección de secretos empresariales con las categorías de riesgo establecidas en normativas internacionales (como la *AI Act* de la UE).

- *Para el sector empresarial (desarrolladores y usuarios):*

- Adopción de auditorías externas independientes: contratar terceros de confianza que verifiquen la equidad del algoritmo bajo acuerdos de no divulgación (NDAs), garantizando la transparencia frente a reguladores sin filtrar datos al mercado.
- Mejora de la documentación técnica: mantener expedientes detallados sobre el origen de los datos de entrenamiento y los procesos de limpieza de sesgos, para facilitar la defensa jurídica del sistema en caso de litigios.
- Comunicación de la lógica: proporcionar a los usuarios finales explicaciones contrafácticas (p. ej.: *si su ingreso hubiera sido de* X*, el crédito habría sido aprobado*) que ofrezcan transparencia sin revelar la

arquitectura interna del modelo.

- *Para el sistema judicial y órganos de control:*

- Creación de peritajes algorítmicos especializados: fomentar la formación de peritos judiciales capaces de evaluar sistemas complejos y actuar como custodios de la información sensible en los procesos legales.
- Criterio de proporcionalidad: aplicar siempre un test de proporcionalidad antes de ordenar la revelación de información confidencial, evaluando si existen medios menos intrusivos para lograr la transparencia requerida.
- Protección de datos personales: asegurar que los esfuerzos de transparencia no comprometan la privacidad de los sujetos cuyos datos alimentan los modelos (privacidad diferencial).

El equilibrio reside en transitar de un modelo de secreto absoluto a uno de opacidad controlada. La protección jurídica del algoritmo es legítima siempre que no se convierta en una patente de corso de la arbitrariedad automatizada. La meta final debe ser una IA que sea innovadora pero, por encima de todo, auditable y justa.

Bibliografía

Aarab, A., El Marzouki, A., Boubker, O., & El Moutaqi, B. (2025). Integrating AI in Public Governance: A Systematic Review. *Digital*, *5*(4), 59. https://doi.org/10.3390/digital5040059

Aguirre Sala, J.F. (2025). La gobernanza de la inteligencia artificial como desafío a los paradigmas del derecho en el siglo XXI. *Revista IUS.* 20(56), 169-193. https://doi.org/10.35487/rius.v20i56.2025.1063

Araya Paz, C. (2021). Transparencia algorítmica ¿un problema normativo o tecnológico?. *CUHSO (Temuco)*. *31*(2), 306-334. https://dx.doi.org/10.7770/cuhso-v31n2-art2196

Azuaje Pirela, M., y Finol González, D. (2020). Transparencia algorítmica y la propiedad intelectual e industrial: tensiones y soluciones. *Revista La Propiedad Inmaterial*. 30, 111–146. https://doi.org/10.18601/16571959.n30.05

Baz Lomba, C. (2021). Los algoritmos y la toma de decisiones administrativas. Especial referencia a la transparencia. *Revista CEFLegal.* 243, 119-160. https://revistas.cef.udima.es/index.php/ceflegal/article/download/9441/9169/17393

Beliz, G. (2025). *Atlas de inteligencia artificial para el desarrollo humano de América Latina y el Caribe.* Nueva York: Programa de las Naciones Unidas para el Desarrollo (PNUD). https://www.undp.org/sites/g/files/zskgke326/files/2025-

06/atlas_a_8_6_compressed_0_0.pdf

Caiza, G., Sanguña, V., Tusa, N., Masaquiza, V., Ortiz, A., & García, MV (2024). Navegando por las opciones gubernamentales: una revisión integral del impacto de la inteligencia artificial en la toma de decisiones. *Informática* , *11* (3), 64. https://doi.org/10.3390/informatics11030064

Carrasco Delgado, B. L. (2025). Justicia y algoritmos: Un análisis ético–jurídico de la Ley 31814 sobre inteligencia artificial en el Perú. *Forseti. Revista De Derecho*, *14*(22), 99–116. https://doi.org/10.21678/forseti.v14i22.2828

Cihon, P., Schuett, J., & Baum, S. D. (2021). Corporate Governance of Artificial Intelligence in the Public Interest. *Information*, *12*(7), 275. https://doi.org/10.3390/info12070275

De Noyette, E., Stähler, L. & Margoni, T. (2025). Data Secrets: The Data Act's New Trade Secrets Framework. *IIC - International Review of Intellectual Property and Competition Law.* 56, 984–1014. https://doi.org/10.1007/s40319-025-01601-9

Delva Benavides, J. E., & Mendez Gonzalez, V. V. (2025). Algoritmos opacos y privacidad diferencial: ¿puede haber transparencia sin comprometer la protección?. *Forseti. Revista De Derecho*. *14*(22), 227–245. https://doi.org/10.21678/forseti.v14i22.2834

Dumouchel, P. (2023). AI and Regulations. *AI*, *4*(4), 1023-1035. https://doi.org/10.3390/ai4040052

Fernandini, C., & Saavedra, Y. (2025). El Derecho de la Innovación: una especialización legal en auge. *Revista De Actualidad Mercantil*, (9), 99–120. https://revistas.pucp.edu.pe/index.php/actualidadmercantil/article/view/30910

Fierro Rodríguez, D. (2024). Gobernanza y Reglamento de Inteligencia Artificial desde la primera óptica de OpenAI. *Derecho & Sociedad*. (63), 243–258. https://doi.org/10.18800/dys.202402.016

Gunasekara, L., El-Haber, N., Nagpal, S., Moraliyage, H., Issadeen, Z., Manic, M., & De Silva, D. (2025). Una revisión sistemática de los principios y prácticas de la inteligencia artificial responsable. *Applied System Innovation*. *8*(4), 97. https://doi.org/10.3390/asi8040097

Ladosky, M. H. G., & López-Martínez, G. (2025). La soledad del *rider.* Normativa y estrategias laborales de repartidores digitales en Brasil y España. *Revista Brasileira De Ciências Sociais*, *40*, e40008. https://doi.org/10.1590/40008/2025

Loayza Villanueva, S. E. (2025). Desafíos y oportunidades jurídicas de la inteligencia artificial en el ámbito laboral: hacia una regulación equitativa y responsable. *Laborem*, *24*(31), pp.139–161. https://doi.org/10.56932/laborem.24.31.6

Mazzinghy, A. O. d. C., Silva, R. M. d. S. e., Fernandes, R. M., Batista, E. D., Picanço, A. R. S., Monteiro, N. J., de Amorim, D. M., Cardoso, B. d. F. O., Silva, J. M. N. d., & Martins, V. W. B. (2025). Assessment of the Benefits of the ISO/IEC 42001 AI Management System: Insights from Selected Brazilian Logistics

Experts: An Empirical Study. *Standards*, *5*(2), 10. https://doi.org/10.3390/standards5020010

Morales Oñate, D. A. (2021). Implicaciones jurídicas del algoritmo: derechos intelectuales y privacidad. *Foro: Revista De Derecho*. *36*, 111-130. https://doi.org/10.32719/26312484.2021.36.6

Mylly, U.M. (2023). Transparent AI? Navigating Between Rules on Trade Secrets and Access to Information. *IIC - International Review of Intellectual Property and Competition Law.* 54, 1013–1043. https://doi.org/10.1007/s40319-023-01328-5

Olugbade, O. (2025). En busca de un mecanismo de gobernanza global para la Inteligencia Artificial (IA): una perspectiva de acción colectiva. *Glob. Public Policy Gov.* 5, 139–161. https://doi.org/10.1007/s43508-025-00113-z

Oncioiu, I., & Bularca, A.R. (2025). Gobernanza de la inteligencia artificial en la educación superior: El papel de las estrategias basadas en el conocimiento para fomentar la conciencia jurídica y la alfabetización ética en inteligencia artificial. *Societies* , *15* (6), 144. https://doi.org/10.3390/soc15060144

Ospina Diaz, M. R., & Zambrano Ospina, K. J. (2023). Gobierno digital e inteligencia artificial, una mirada al caso colombiano. *Administración & Desarrollo*. *53*(1), 1-34. https://doi.org/10.22431/25005227.vol53n1.2

Ozmen Garibay, O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C., … Xu, W. (2023). Seis grandes desafíos de la inteligencia artificial centrada en el ser humano. *International Journal of Human–Computer*

Interaction , *39* (3), 391–437. https://doi.org/10.1080/10447318.2022.2153320

Ruiz Tarrías, S. (2023). La búsqueda de un modelo regulatorio de la ia en la Unión Europea. *Anales De La Cátedra Francisco Suárez*. *57*, 91–119. https://doi.org/10.30827/acfs.v57i.25245

Sklavos, G., Theodossiou, G., Papanikolaou, Z., Karelakis, C., & Ragazou, K. (2024). Gobernanza de inteligencia artificial basada en criterios ambientales, sociales y de gobernanza: digitalización del liderazgo y la gestión de recursos humanos de las empresas. *Sustainability* , *16* (16), 7154. https://doi.org/10.3390/su16167154

The Intelligence. (s/f). La justicia respalda la transparencia obligatoria en inteligencia artificial y abre una nueva etapa regulatoria. The Intelligence. https://theintelligence.es/la-justicia-respalda-la-transparencia-obligatoria-en-inteligencia-artificial-y-abre-una-nueva-etapa-regulatoria/

Toche, F. (s/f). Protección de Datos Personales y Decisiones Automatizadas: Sentencia del Tribunal de Justicia de la Unión Europea (TJUE) – Asunto C-203/22. *IALaw*. https://www.iriartelaw.com/2025/03/20/proteccion-de-datos-personales-y-decisiones-automatizadas-sentencia-del-tribunal-de-justicia-de-la-union-europea-tjue-asunto-c-203-22/

Viveros Zuazo, A. A. (2015). El Riesgo Moral y la Regulación de la Calidad de los Servicios Públicos. *Derecho & Sociedad*. (45), 45–52. https://revistas.pucp.edu.pe/index.php/derechoysociedad/article/view/15223

De esta edición de “*Protección de secretos empresariales vs. transparencia algorítmica”*, se terminó de editar en la ciudad de Colonia del Sacramento en la República Oriental del Uruguay el 18 de febrero de 2026

EST. 2021 EMC
EDITORIAL MAR CARIBE

PROTECCIÓN DE SECRETOS EMPRESARIALES VS. TRANSPARENCIA ALGORÍTMICA

ESCRITO POR:

SANTIAGO GONZALES MESIA
ROBERTO SEGUNDO TEJADA RODRIGUEZ
JOSÉ LUIS CASTRO ULLILEN
VIVIANA INÉS VELLÓN FLORES DE SOLANO
TIMOTEO SOLANO ARMAS
CARLOS MÁXIMO GONZÁLES AÑORGA

ISBN: 978-9915-698-63-2
9 789915 698632