EDUCA - PLUS
EDITORIAL ACADÉMICA

DERECHO E INTELIGENCIA ARTIFICIAL



Octubre 2026 – Educa Plus Editorial Académica
Copyright © Educa Plus Editorial Académica
Copyright del texto © 2026 de Autores
Ecuador, Naranjito, Cod. Post. 091803
Email: editorial@educapluseditorial.com
Contacto: +593 960 838 709
https://biblioteca.educapluseditorial.com/

Datos Técnicos de Publicación Internacional
Título: Derecho e inteligencia Artificial **Subtítulo:** Regulación, responsabilidad, ética y derechos, fundamentales en la era algorítmica en el contexto educativo peruano **Autores:** Guerra Rodríguez, L. A., & Leyton Cotrina, B. M. **Diseño y Producción Editorial:** Durán Sánchez, Stephany **Editor:** Barco Acosta, Jonathan **Formato:** PDF **Páginas:** 245 pág. **Tamaño:** A5 14,8 x 21 cm **Modo de acceso:** World Wide Web **ISBN:** 978-9907-9622-0-8 **DOI:** https://doi.org/10.66646/educa-plus/10oct2619 **Prefijo Editorial ISBN:** 978-9907-9571

Este libro fue sometido a arbitraje académico bajo el sistema de evaluación por pares doble ciego (double-blind peer review).

Disponible para su descarga gratuita en https://biblioteca.educapluseditorial.com/

EDITORIAL

Los contenidos de este libro pueden ser descargados, reproducidos, difundidos e impresos con fines de estudio, investigación y docencia, o para su utilización en productos o servicios no comerciales, siempre que se reconozca adecuadamente a los autores como fuente y titulares de los derechos de propiedad intelectual, sin que ello implique en modo alguno que aprueban las opiniones, productos o servicios resultantes. En el caso de contenidos que indiquen expresamente que proceden de terceros, deberán dirigirse a la fuente original indicada para gestionar los permisos correspondientes.

Cómo citar esta obra

Guerra Rodríguez, L. A., & Leyton Cotrina, B. M. (2026). Derecho e inteligencia artificial: Regulación, responsabilidad, ética y derechos fundamentales en la era algorítmica. Educa Plus Editorial Académica. https://doi.org/10.66646/educa-plus/10oct2619

Proceso Editorial:

Cada uno de los textos publicados por **Educa Plus Editorial Académica** ha sido sometido a un riguroso proceso de **evaluación por pares doble ciego externos** (*double-blind peer review*), conforme a los estándares editoriales internacionales y a la normativa interna de la editorial.

Aviso Legal:

La información presentada, así como el contenido, fotografías, gráficos, cuadros, tablas y referencias de este manuscrito, es de exclusiva responsabilidad del/los autor/es y no refleja necesariamente el pensamiento ni la posición institucional de la **Educa Plus Editorial Académica**. Las opiniones expresadas en cada capítulo son responsabilidad exclusiva de sus respectivos autores.

Derechos de Autor ©

Este documento se publica bajo los términos y condiciones de la licencia:

Creative Commons Reconocimiento — No Comercial — Compartir Igual 4.0 Internacional (CC BY-NC-SA 4.0)

Usted es libre de compartir y adaptar el material, siempre que otorgue el crédito correspondiente, no lo utilice con fines comerciales y distribuya sus contribuciones bajo la misma licencia.

Para ver una copia de esta licencia, visite: https://creativecommons.org/licenses/by-nc-sa/4.0/deed.es



DEDICATORIA

Dedico esta obra, con profundo amor y gratitud, a mi madre, **Carol Katya Rodríguez Muñoz, viuda de Guerra**, quien me enseñó que la fe en Dios, el esfuerzo y la perseverancia permiten alcanzar nuestros sueños. Su fortaleza, sacrificio y amor incondicional han sido mi mayor inspiración.

A mis hermanitos **Luis Jesús, Josécito y Luciana Milagros**, mis angelitos y estrellas que iluminan mi camino. A mis familiares y amigos, por su cariño, confianza y apoyo en los momentos más importantes de mi vida.

A la memoria de mis abuelos maternos, Orfelinda Muñoz Rodríguez de Rodríguez y Jerónimo Nicanor Rodríguez Ruiz, y especialmente de mi amado padre, **José Luis Guerra Lu**, cuya ausencia física jamás borrará su presencia en mi corazón.

Papá, tus enseñanzas, tus recuerdos y el amor que me brindaste siguen acompañándome en cada paso. Este libro representa horas de esfuerzo, aprendizaje y perseverancia, pero también un homenaje a tu memoria. Cada logro alcanzado lleva una parte de ti, porque tu ejemplo continúa guiando mis sueños y fortaleciendo mi camino profesional.

Con humildad, amor y eterna gratitud: **¡Gracias, gracias, gracias!**

Dr. Luis Antonio Guerra Rodríguez

Dedico estas páginas, con profundo amor y gratitud, a mi amada madre, **Míriam**, quien ha sido el pilar fundamental de mi vida y una fuente inagotable de fortaleza, inspiración y esperanza.

Esta obra representa años de esfuerzo, aprendizaje, dedicación y compromiso con mi crecimiento profesional. Cada desafío superado y cada conocimiento adquirido han fortalecido mi vocación y mi deseo de contribuir, mediante mi trabajo, al desarrollo de la sociedad.

Que estas páginas sean testimonio de perseverancia, responsabilidad y pasión por el conocimiento, así como un reflejo del compromiso de seguir creciendo, aprendiendo y dejando una huella positiva en mi profesión.

Dr. Brayhan Martín Leyton Cotrina

PRESENTACIÓN

La inteligencia artificial ya participa en actividades que inciden en la vida cotidiana, las instituciones y el ejercicio de los derechos. Su incorporación plantea preguntas jurídicas que no pueden responderse únicamente desde la técnica: quién define sus fines, qué información procesa, cómo influye en una decisión y quién debe responder cuando produce un efecto adverso. Este libro aborda esas preguntas desde una perspectiva jurídica, interdisciplinaria y comparada, con atención a los debates internacionales y latinoamericanos.

Derecho e inteligencia artificial: Regulación, responsabilidad, ética y derechos fundamentales en la era algorítmica organiza el análisis en diez capítulos. Parte de los conceptos y principios éticos que permiten comprender los sistemas de IA, y examina después su relación con la dignidad, la igualdad, la libertad, la privacidad y la protección de datos. Compara los principales marcos regulatorios internacionales y latinoamericanos; estudia la responsabilidad por daños; y analiza usos de IA en la administración pública, la justicia y la actividad empresarial, junto con los retos de propiedad intelectual y contratación.

El capítulo final integra estos temas y considera cómo fortalecer la cooperación internacional, la supervisión y las herramientas de evaluación. A lo largo de la obra, la atención se dirige tanto a las capacidades de la tecnología como a las instituciones y relaciones sociales en que se utiliza. La automatización puede contribuir a mejorar servicios y ampliar capacidades, pero sus resultados requieren evaluación contextual, responsabilidades identificables y vías efectivas de revisión y reparación.

La obra está dirigida a estudiantes y profesionales del Derecho, investigadores, docentes, responsables de políticas públicas y personas interesadas en la relación entre tecnología y sociedad. Su propósito es ofrecer bases conceptuales y jurídicas para un análisis riguroso, promover una discusión informada y contribuir a que el desarrollo de la inteligencia artificial sea compatible con los derechos fundamentales y el Estado de derecho.

CONTENIDO

Presentación 6

Capítulo 1
Inteligencia artificial y derecho: fundamentos, conceptos y transformación del orden jurídico 12

Introducción 13
1.1. Concepto, tipos y funcionamiento de los sistemas de inteligencia artificial 14
1.2. La era algorítmica y el impacto de la IA en las instituciones jurídicas 20
1.3. Derecho y tecnología: tensiones entre innovación, regulación y seguridad jurídica 24
1.4. Enfoques metodológicos y teóricos para el estudio jurídico de la IA 28

Capítulo 2
Fundamentos éticos y principios de la inteligencia artificial 34

Introducción 35
2.1. Ética de la IA: principios internacionales y corrientes teóricas 36
2.2. Transparencia, explicabilidad y rendición de cuentas 42
2.3. Sesgos algorítmicos, equidad y no discriminación 46
2.4. De la ética al derecho: soft law, autorregulación y su fuerza normativa 50

Capítulo 3
Derechos fundamentales en la era algorítmica 56

Introducción 57
3.1. Dignidad humana, autonomía y libertad frente a los sistemas automatizados 58
3.2. Igualdad, no discriminación y justicia algorítmica 62

3.3. Libertad de expresión, información y pluralismo en entornos algorítmicos 66
3.4. Derechos laborales, sociales y de grupos vulnerables ante la IA 70

Capítulo 4
Protección de datos personales, privacidad y decisiones automatizadas 76

Introducción 77
4.1. Datos personales, privacidad y autodeterminación informativa 78
4.2. Elaboración de perfiles y decisiones individuales automatizadas 83
4.3. Biometría, reconocimiento facial y vigilancia algorítmica 88
4.4. Derechos de los titulares y mecanismos de tutela en entornos de IA 92

Capítulo 5
Regulación internacional y comparada de la inteligencia artificial 98

Introducción 99
5.1. El Reglamento de IA de la Unión Europea y el enfoque basado en riesgos 100
5.2. Modelos regulatorios en Estados Unidos, China y otras jurisdicciones 104
5.3. Instrumentos de organismos internacionales: UNESCO, OCDE, ONU y Consejo de Europa 111
5.4. El efecto Bruselas, la armonización normativa y la competencia regulatoria 114

Capítulo 6
Regulación de la inteligencia artificial en América Latina 118

Introducción 119
6.1. Panorama normativo y estrategias nacionales de IA en la región 120
6.2. Proyectos de ley y experiencias regulatorias en países latinoamericanos 125
6.3. Sistema interamericano de derechos humanos y estándares

aplicables a la IA 132
6.4. Desafíos de implementación, capacidades institucionales y cooperación regional 138

Capítulo 7
Responsabilidad jurídica por daños causados por sistemas de IA 144

Introducción 145
7.1. Responsabilidad civil: culpa, riesgo y responsabilidad por producto defectuoso 146
7.2. Atribución de responsabilidad entre desarrolladores, proveedores, desplegadores y usuarios 151
7.3. Responsabilidad penal, administrativa y sancionatoria 157
7.4. Seguros, carga de la prueba y mecanismos de reparación 161

Capítulo 8
Inteligencia artificial en la administración pública y la justicia168

8.1. Decisiones administrativas automatizadas y debido proceso algorítmico 170
8.2. IA en la administración de justicia: justicia predictiva y apoyo judicial 174
8.3. IA en seguridad, policía predictiva y control penal 179
8.4. Control judicial, transparencia y límites del uso estatal de la IA 185

Capítulo 9
Propiedad intelectual, contratación y actividad empresarial con IA 190

9.1. Autoría, titularidad y protección de obras y creaciones generadas por IA 192
9.2. Datos de entrenamiento, derechos de autor y uso legítimo 197
9.3. Contratos inteligentes, contratación tecnológica y cumplimiento normativo 202

9.4. Competencia, mercados digitales y gobernanza corporativa de la IA 207

Capítulo 10

Gobernanza global de la IA y perspectivas futuras del derecho 214

10.1. Gobernanza multinivel y cooperación internacional 216
10.2. IA generativa, sistemas autónomos y nuevos desafíos regulatorios 221
10.3. Hacia un marco jurídico adaptativo: sandboxes, auditorías y evaluaciones de impacto 226
10.4. Recomendaciones y agenda de política pública para una IA centrada en las personas 231

Bibliografía 238

Capítulo 1

Inteligencia artificial y derecho: fundamentos, conceptos y transformación del orden jurídico



INTRODUCCIÓN

La relación entre inteligencia artificial y derecho exige aclarar primero qué se entiende por sistema de inteligencia artificial y qué aspectos de su funcionamiento resultan jurídicamente relevantes. La expresión "inteligencia artificial" comprende técnicas diversas, desde sistemas que clasifican datos hasta modelos capaces de generar textos, imágenes o recomendaciones. Por ello, una regulación basada únicamente en el nombre de una tecnología puede quedar desactualizada o abarcar actividades muy diferentes. El análisis jurídico requiere observar qué sistema se utiliza, quién lo desarrolla o despliega, qué función cumple, en qué contexto opera y qué efectos produce sobre personas, instituciones y bienes.

Este capítulo establece el marco conceptual de la obra. Primero delimita el concepto de sistema de inteligencia artificial y presenta sus principales tipos y componentes; después examina la influencia de la era algorítmica sobre las instituciones jurídicas. A continuación, aborda las tensiones entre innovación, regulación y seguridad jurídica, y concluye con los enfoques metodológicos y teóricos apropiados para estudiar la inteligencia artificial desde el derecho. La exposición distingue entre descripciones técnicas, categorías analíticas y normas vigentes, pues cada plano responde a preguntas distintas. La finalidad es contar con conceptos consistentes para analizar, en los capítulos posteriores, los derechos fundamentales, la protección de datos, la regulación comparada y la responsabilidad jurídica.

1.1. Concepto, tipos y funcionamiento de los sistemas de inteligencia artificial

1.1.1. Delimitación conceptual y relevancia jurídica

No existe una definición única de inteligencia artificial que satisfaga todos los propósitos científicos, técnicos y jurídicos. En el campo de la informática, el término suele describir sistemas computacionales capaces de realizar tareas asociadas con capacidades como la percepción, el aprendizaje, la predicción, el razonamiento o la generación de contenido. Sin embargo, esta caracterización puede ser demasiado amplia para regular conductas concretas: algunas herramientas automatizan operaciones simples, mientras que otras identifican patrones en grandes conjuntos de datos y producen resultados que influyen en decisiones humanas o institucionales.

La definición de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), revisada para reflejar avances recientes, ofrece una referencia funcional. Describe el sistema de IA como un sistema basado en máquinas que, para objetivos explícitos o implícitos, infiere a partir de las entradas recibidas cómo generar resultados —como predicciones, contenido, recomendaciones o decisiones— que pueden influir en entornos físicos o virtuales. Añade que los sistemas difieren en sus niveles de autonomía y adaptabilidad después de su implementación (OCDE, 2024). Esta formulación se concentra en capacidades y resultados observables y no atribuye al sistema conciencia, intención moral ni personalidad jurídica.

El Reglamento de Inteligencia Artificial de la Unión Europea adopta una definición normativa próxima a ese enfoque. Su artículo 3 considera sistema de IA aquel basado en máquinas, diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación después de su implementación, que infiere cómo generar, a partir de las entradas recibidas, resultados como predicciones, contenidos, recomendaciones o decisiones capaces de

influir en entornos físicos o virtuales (Reglamento (UE) 2024/1689, art. 3.1). La definición es aplicable en el ámbito propio de ese reglamento y no constituye automáticamente una norma universal para los países latinoamericanos. Su utilidad comparada radica en que identifica elementos relevantes para el análisis: la inferencia, los resultados, la autonomía, la adaptabilidad y la influencia sobre el entorno.

Desde el derecho, la pregunta central no es si una máquina "piensa" en sentido humano, sino qué actividad se automatiza y qué efectos jurídicos se derivan de su uso. Una herramienta puede carecer de comprensión humana y, aun así, producir consecuencias importantes: ordenar solicitudes de crédito, recomendar una contratación, estimar la probabilidad de reincidencia o priorizar casos en un servicio público. La capacidad técnica no determina por sí sola la legitimidad jurídica. Esta depende también de la finalidad, la calidad de los datos, las condiciones de uso, la distribución de responsabilidades y las posibilidades de supervisión y revisión.

Por la misma razón, conviene distinguir entre inteligencia artificial como campo de investigación, sistema de IA como objeto técnico y decisión automatizada como proceso decisorio en el que interviene una herramienta computacional. Un sistema puede asistir a una persona sin adoptar la decisión final; también puede producir una recomendación que, en la práctica, determine el resultado porque la persona encargada carece de tiempo, formación o autoridad para apartarse de ella. En el segundo supuesto, la intervención humana podría existir formalmente, pero ser insuficiente en términos sustantivos. El análisis jurídico debe observar cómo opera realmente el proceso y no limitarse a la descripción institucional de que "la decisión la toma un funcionario".

1.1.2. Tipos, componentes y ciclo de funcionamiento

La clasificación de los sistemas de IA puede realizarse desde distintos criterios. Según la técnica empleada, se encuentran sistemas basados en reglas, aprendizaje automático, aprendizaje profundo, procesamiento

del lenguaje natural, visión computacional y modelos generativos. Estas categorías pueden superponerse: un producto puede combinar redes neuronales, procesamiento del lenguaje y reglas programadas. Por ello, la clasificación sirve para describir funciones y riesgos, pero no siempre permite establecer por sí sola el régimen jurídico aplicable.

Los sistemas basados en reglas ejecutan instrucciones previamente definidas por personas. Cuando se cumplen ciertas condiciones, el programa produce una respuesta prevista. Su comportamiento suele ser más rastreable que el de algunos modelos estadísticos, aunque esto no elimina posibles errores de diseño ni sesgos incorporados en las reglas. En cambio, el aprendizaje automático identifica patrones a partir de datos y ajusta parámetros para mejorar una tarea conforme a un criterio de optimización. El aprendizaje profundo emplea redes neuronales con múltiples capas para detectar relaciones complejas, por ejemplo, en imágenes, audio o lenguaje. Los modelos generativos producen nuevos contenidos a partir de patrones aprendidos en datos de entrenamiento y de instrucciones o entradas proporcionadas durante su uso.

Un sistema de IA suele integrar varios componentes. Recibe datos o instrucciones; los procesa mediante un modelo y una arquitectura computacional; genera una salida; y se incorpora a un entorno de uso en el que esa salida puede orientar una acción. En determinados sistemas también hay retroalimentación, pues los resultados posteriores influyen en ajustes o versiones futuras. No todos los sistemas aprenden continuamente durante su operación: algunos permanecen estáticos después de entrenarse, mientras otros pueden adaptarse mediante nuevos datos o actualizaciones. Esta diferencia importa para la auditoría, la atribución causal y la conservación de evidencias.

El ciclo de vida suele comprender las etapas de planificación y diseño, recopilación y preparación de datos, entrenamiento o configuración del modelo, verificación y validación, despliegue, operación, supervisión y eventual modificación o retirada. En cada etapa pueden intervenir

actores distintos. El desarrollador puede no controlar la decisión organizacional para la que el sistema se usa; el proveedor puede integrar un modelo de terceros; y la entidad usuaria puede aportar datos o adaptar la herramienta al contexto. La cadena de participación ayuda a identificar responsabilidades y obligaciones de diligencia, pero su distribución concreta depende del ordenamiento aplicable y de las circunstancias del caso.

Los resultados tampoco son homogéneos. Un sistema puede clasificar, detectar, predecir, recomendar, generar contenido o decidir. La diferencia funcional es importante: la generación de una recomendación no equivale necesariamente a la adopción de una decisión, aunque la recomendación pueda tener un efecto determinante. Además, los resultados pueden ser probabilísticos, por lo que expresan estimaciones sujetas a error y no certezas sobre personas o acontecimientos. Presentar una puntuación de riesgo como si fuera una verdad objetiva confunde un resultado computacional con el juicio jurídico y puede distorsionar la valoración de la evidencia.

En materia jurídica conviene describir los sistemas por su función, contexto, nivel de intervención y consecuencias previsibles. Una herramienta que sugiere rutas de lectura en una biblioteca no plantea los mismos problemas que otra que influye en el acceso a prestaciones, empleo, crédito o justicia. El riesgo no depende exclusivamente de la complejidad técnica. También importan la vulnerabilidad de las personas afectadas, la magnitud del posible daño, la posibilidad de corregir un error y la existencia de mecanismos efectivos de revisión.

Esta aproximación explica el valor de los marcos regulatorios que combinan definiciones funcionales con obligaciones ajustadas a la actividad y al nivel de riesgo. La Recomendación de la UNESCO sobre la Ética de la Inteligencia Artificial, adoptada en 2021, es un instrumento internacional de orientación ética y de política pública; promueve un enfoque centrado en los derechos humanos, la dignidad y

la supervisión durante el ciclo de vida, pero no equivale por sí misma a un tratado vinculante (UNESCO, 2021). La Convención Marco del Consejo de Europa sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho, abierta a la firma en 2024, sí es un tratado internacional; su existencia y alcance deben distinguirse de su entrada en vigor para cada Estado conforme a los requisitos correspondientes.

Tabla 1.1
Tipos de sistemas de inteligencia artificial y relevancia jurídica

Tipo de sistema	Función principal	Ejemplo de aplicación	Pregunta jurídica relevante
Basado en reglas	Ejecuta instrucciones previamente definidas por personas	Verificar si una solicitud cumple requisitos establecidos	¿Las reglas son claras, legales y aplicadas de manera uniforme?
Aprendizaje automático	Identifica patrones en datos para clasificar o predecir resultados	Estimar el riesgo de impago de una solicitud de crédito	¿Los datos son pertinentes y representativos? ¿El sistema produce impactos desiguales?
Aprendizaje profundo	Procesa patrones complejos mediante redes neuronales de múltiples capas	Reconocer objetos en imágenes o analizar audio	¿Se han evaluado sus errores y limitaciones en el contexto de uso?
Procesamiento del lenguaje natural	Analiza o produce lenguaje humano	Clasificar consultas, resumir documentos o responder preguntas	¿La salida puede contener errores, omisiones o información confidencial?

Visión computacional	Analiza imágenes o videos para detectar, clasificar o identificar elementos	Reconocimiento facial o inspección automatizada de imágenes	¿El tratamiento es necesario y proporcional? ¿Cómo se controlan errores de identificación?
Generativo	Produce textos, imágenes, audio, video o código a partir de instrucciones y patrones aprendidos	Generar un borrador de texto o una imagen	¿Qué revisión humana requiere el resultado y qué derechos pueden afectar los datos de entrada o la salida?

Nota. *Las categorías pueden superponerse: un mismo producto puede combinar varias técnicas. La clasificación describe capacidades generales; por sí sola no determina el nivel de riesgo ni el régimen jurídico aplicable.*

Figura 1.1.

Ciclo de vida de un sistema de inteligencia artificial y puntos de control jurídico



Nota: *Elaboración propia. No atribuir autonomía jurídica al sistema; mostrar los actores humanos y organizacionales en cada etapa.*

1.2. La era algorítmica y el impacto de la IA en las instituciones jurídicas

1.2.1. La mediación algorítmica de la vida social

La expresión "era algorítmica" alude a una organización social en la que procedimientos computacionales participan de manera creciente en la selección, clasificación, priorización y distribución de información, servicios y oportunidades. Los algoritmos no son necesariamente sistemas de IA: un algoritmo puede consistir en instrucciones deterministas que no aprenden ni generan inferencias estadísticas. Sin embargo, en el debate público suelen mezclarse ambos conceptos porque muchas aplicaciones contemporáneas combinan reglas, modelos de aprendizaje y sistemas de recomendación.

La relevancia jurídica de esta mediación aparece cuando los sistemas influyen en decisiones que afectan intereses protegidos. Una plataforma puede ordenar contenidos y, con ello, incidir en la visibilidad de ideas; un empleador puede usar herramientas de selección para filtrar candidaturas; una entidad financiera puede emplear modelos de riesgo para valorar solicitudes; y una administración puede automatizar la gestión de expedientes. En cada caso, el sistema reconfigura el modo en que se ejerce un poder, aunque no sustituya formalmente a la institución que lo detenta.

El uso de herramientas algorítmicas también modifica la escala y velocidad de las actuaciones. Una persona puede evaluar algunos expedientes en un tiempo limitado; un sistema puede procesar miles de registros en segundos. Esta capacidad amplía la cobertura de ciertos servicios, pero también facilita que errores o criterios discutibles se apliquen de forma masiva. Cuando la automatización se integra en procedimientos esenciales, una falla técnica, una variable inadecuada o un dato incorrecto puede afectar simultáneamente a numerosas personas. La magnitud del impacto exige evaluar no solo cada decisión individual, sino también los patrones colectivos que se producen.

La mediación algorítmica puede hacer menos visible quién decidió y sobre qué base. En un procedimiento tradicional, la decisión suele atribuirse a una autoridad identificable que expresa razones. En un proceso asistido por modelos, pueden intervenir el proveedor de software, quienes seleccionaron los datos, la entidad que definió la finalidad, el equipo que configuró el umbral y la persona que adoptó el resultado. La fragmentación de la cadena puede provocar una "zona de irresponsabilidad" si cada participante afirma que la decisión pertenece a otro. La organización debe conservar trazabilidad suficiente para reconstruir el proceso, identificar intervenciones y determinar quién tenía capacidad real para prevenir o corregir el daño.

La opacidad puede ser técnica, organizacional o institucional. Es técnica cuando el funcionamiento del modelo resulta difícil de interpretar; organizacional cuando la documentación y las decisiones están distribuidas entre equipos o proveedores; e institucional cuando la persona afectada desconoce la existencia del sistema, la forma de impugnarlo o la autoridad responsable. Una explicación útil no exige siempre revelar cada parámetro del modelo. Debe permitir comprender los factores relevantes, el papel que tuvo la herramienta, la finalidad perseguida, las limitaciones conocidas y las vías disponibles para cuestionar el resultado.

La automatización también plantea una cuestión de autoridad epistémica: la tendencia a considerar que la salida del sistema es objetiva por haber sido producida mediante cálculos. Sin embargo, los modelos incorporan decisiones humanas sobre la definición del problema, la selección de datos, las variables, los umbrales de error aceptables y la finalidad de optimización. La cuantificación puede hacer más uniforme una práctica, pero no la vuelve neutral. La institución debe explicar por qué la variable escogida es pertinente y por qué el resultado es suficiente para justificar la medida adoptada.

1.2.2. Transformaciones de las instituciones y las prácticas jurídicas

Las instituciones jurídicas utilizan tecnologías digitales desde hace décadas, por ejemplo, en la gestión documental, la búsqueda de precedentes y la administración de expedientes. La incorporación de IA amplía esas capacidades al permitir clasificar documentos, resumir información, detectar relaciones, traducir textos o generar borradores. Estas aplicaciones pueden facilitar tareas rutinarias y mejorar el acceso a información, siempre que exista validación humana y que la herramienta no se confunda con una fuente jurídica ni con un razonamiento autorizado.

En el ámbito de la abogacía, los sistemas generativos pueden asistir en la preparación de documentos, la organización de expedientes o la identificación preliminar de cuestiones jurídicas. El profesional sigue obligado a verificar la exactitud, la vigencia y la pertinencia de las fuentes, así como a proteger la confidencialidad de la información. La fluidez de una respuesta generada no demuestra que sus citas existan ni que su interpretación sea correcta. La responsabilidad profesional se conserva, aunque una parte de la tarea se delegue en una herramienta.

En la administración pública, la IA puede emplearse para gestionar solicitudes, detectar inconsistencias, priorizar inspecciones o apoyar la prestación de servicios. El posible beneficio de mayor rapidez debe evaluarse junto con las garantías de legalidad, motivación, igualdad y revisión. La autoridad pública no puede eludir sus deberes atribuyendo el resultado a un proveedor o a un modelo. Si el sistema contribuye a una decisión que restringe derechos o asigna beneficios, la entidad debe poder explicar la base jurídica, el rol del sistema y el procedimiento de impugnación. La automatización no transforma una competencia pública en una operación puramente técnica.

La administración de justicia presenta retos particulares porque el uso de IA puede afectar la independencia, la imparcialidad, la igualdad

procesal y la confianza pública. Herramientas de búsqueda o gestión pueden apoyar labores auxiliares, pero los sistemas que estiman la conducta de personas, proponen resultados o influyen en la valoración de riesgos requieren cautela reforzada. El juez o funcionario debe conservar control efectivo y autonomía de criterio. Una salida estadística no reemplaza la valoración individual de hechos y pruebas, ni puede operar como presunción de culpabilidad o como fundamento secreto de una decisión.

La práctica legislativa también se transforma. Los sistemas pueden ayudar a comparar textos, detectar inconsistencias y analizar comentarios públicos; sin embargo, la formulación de normas requiere deliberación democrática, consideración de valores y responsabilidad política. El uso de herramientas de generación de texto puede acelerar borradores, pero no sustituye la evaluación de impactos, la consulta a sectores afectados ni la coherencia con el ordenamiento. La decisión sobre qué riesgos son aceptables y qué derechos deben prevalecer sigue siendo normativa y debe atribuirse a órganos legitimados.

La era algorítmica incide igualmente en la producción de evidencia. Registros de uso, versiones de modelos, instrucciones de entrada, parámetros de configuración y trazas pueden resultar esenciales para conocer cómo se generó una salida. Si esos elementos no se conservan, puede ser difícil comprobar un error o establecer la relación causal entre el sistema y el daño. Las reglas procesales deberán ofrecer mecanismos para obtener información relevante sin desconocer secretos comerciales legítimos, protección de datos o seguridad informática. El equilibrio requiere procedimientos específicos de acceso, peritaje y contradicción, no una exclusión general de la evidencia técnica.

Por último, la transformación no es exclusivamente tecnológica. También es organizacional y cultural. La adopción de IA modifica rutinas, distribuye tareas, define indicadores de desempeño y puede alterar la relación entre la institución y las personas usuarias. El derecho

debe observar la estructura completa en la que opera el sistema: quién lo adquiere, quién lo supervisa, qué capacitación recibe el personal, qué canales existen para reclamar y cómo se evalúan sus resultados. Sin estas condiciones, incluso un sistema técnicamente sofisticado puede agravar desigualdades o debilitar la confianza institucional.

1.3. Derecho y tecnología: tensiones entre innovación, regulación y seguridad jurídica

1.3.1. Innovación, incertidumbre y función reguladora

La innovación tecnológica crea oportunidades y riesgos que no siempre pueden anticiparse con precisión. La IA puede apoyar diagnósticos, mejorar la accesibilidad, optimizar procesos productivos y facilitar el análisis de información. También puede ampliar prácticas de vigilancia, reproducir discriminación, intensificar la concentración económica y dificultar la identificación de responsabilidades. La valoración jurídica debe evitar tanto la aceptación automática de toda innovación como su rechazo general. Debe preguntarse qué problema resuelve la herramienta, quién obtiene el beneficio, quién soporta los riesgos y qué garantías hacen aceptable su uso.

La regulación cumple varias funciones. Establece límites a conductas incompatibles con derechos y principios constitucionales; distribuye deberes entre los actores; crea procedimientos de evaluación, documentación y supervisión; y ofrece remedios cuando se producen daños. También reduce incertidumbre al establecer reglas comunes y estándares de diligencia. Un marco claro puede promover innovación responsable, pues permite a organizaciones y usuarios conocer de antemano las condiciones para diseñar, adquirir o desplegar sistemas.

El desafío está en regular una tecnología que cambia sin fijar reglas excesivamente rígidas. Una norma que enumera productos o técnicas específicas puede quedar obsoleta; una norma que emplea conceptos demasiado abstractos puede generar aplicación inconsistente. Los

enfoques funcionales y basados en riesgos procuran responder a este dilema: definen los sistemas por sus capacidades y adaptan las obligaciones al contexto de uso, al tipo de impacto y a la gravedad del daño potencial. No obstante, la aplicación de esos enfoques depende de criterios suficientemente claros, instituciones competentes y mecanismos de revisión.

La seguridad jurídica exige previsibilidad, coherencia y posibilidad de conocer las consecuencias jurídicas de una conducta. En materia de IA, también requiere que las personas puedan saber cuándo un sistema influye en decisiones relevantes, qué entidad responde por su uso y cómo ejercer sus derechos. La previsibilidad no significa garantizar que todo resultado algorítmico pueda anticiparse. Significa establecer procedimientos para manejar la incertidumbre, comunicar límites, auditar comportamientos y corregir fallos.

Una tensión central surge entre transparencia y protección de información legítimamente reservada. La divulgación completa del código fuente no siempre permite comprender un modelo y puede comprometer la seguridad o derechos de terceros. A la vez, invocar el secreto comercial no debe impedir que una persona conozca las razones esenciales de una decisión que la afecta ni que una autoridad supervise el cumplimiento normativo. La transparencia debe definirse según la finalidad: documentación técnica para auditores; información comprensible para personas afectadas; y datos suficientes para jueces o autoridades competentes. Cada nivel responde a necesidades distintas.

Otra tensión se produce entre eficiencia y garantías procesales. La automatización puede reducir tiempos, pero la rapidez no justifica procedimientos que impidan ser oído, corregir información inexacta o cuestionar un resultado. Del mismo modo, la precisión estadística promedio no determina si la decisión es justa en un caso particular. La legalidad requiere valorar la pertinencia de la herramienta y establecer salvaguardas proporcionales al tipo de decisión, especialmente cuando

están comprometidos derechos fundamentales.

1.3.2. Enfoques regulatorios y criterios para evaluar la seguridad jurídica

La regulación de la IA suele combinar normas generales y reglas sectoriales. Los principios constitucionales, las normas de protección de datos, el derecho del consumidor, las reglas laborales, las normas antidiscriminatorias y los procedimientos administrativos continúan siendo aplicables cuando una tecnología participa en una actividad regulada. La IA no crea un espacio jurídico vacío. Sin embargo, la existencia de normas generales no siempre resuelve cuestiones específicas como la documentación del modelo, la evaluación de riesgos, la supervisión humana o el acceso a información técnica.

Un modelo regulatorio puede ser horizontal, si establece obligaciones generales para sistemas de distintos sectores, o sectorial, si regula su uso en ámbitos como salud, finanzas, educación, empleo o justicia. El enfoque horizontal puede facilitar coherencia y evitar vacíos, aunque corre el riesgo de imponer exigencias uniformes a actividades muy diferentes. Las reglas sectoriales permiten atender características específicas, pero pueden generar fragmentación y duplicidades. Una arquitectura equilibrada requiere coordinación, definiciones compatibles y claridad sobre qué autoridad supervisa cada obligación.

El enfoque basado en riesgos diferencia los sistemas según la probabilidad y gravedad de los impactos previsibles. El Reglamento (UE) 2024/1689 constituye un ejemplo legislativo de aplicación regional que combina prohibiciones para ciertas prácticas, requisitos para determinadas categorías de alto riesgo y obligaciones de transparencia para usos específicos, además de reglas sobre modelos de propósito general. Su ámbito y sus obligaciones se determinan conforme a sus disposiciones; no debe presentarse como una ley universal. Para los países fuera de la Unión Europea, su relevancia puede ser comparativa o práctica cuando una entidad opera en mercados sujetos a ese marco,

pero no sustituye la legislación nacional aplicable.

La regulación también puede apoyarse en instrumentos de derecho blando, como recomendaciones, estándares técnicos y códigos de conducta. Estos instrumentos no suelen tener la misma fuerza vinculante que una ley o un tratado, pero pueden orientar políticas públicas, prácticas organizacionales y futuras reformas. La Recomendación de la UNESCO de 2021 es un estándar global de ética y gobernanza, mientras que la Convención Marco del Consejo de Europa es un tratado que crea obligaciones internacionales para los Estados parte conforme a su vigencia y aplicación. Distinguir los instrumentos evita atribuirles efectos jurídicos que no poseen y permite comprender cómo los principios éticos pueden influir progresivamente en normas vinculantes.

La seguridad jurídica no se agota en la claridad de las definiciones. También depende de la capacidad de las instituciones para aplicar reglas y ofrecer remedios. Una legislación técnicamente avanzada puede resultar ineficaz si las autoridades carecen de personal especializado, presupuesto o acceso a peritajes independientes. Del mismo modo, la supervisión puede perder eficacia si depende exclusivamente de información entregada por el proveedor o si las personas afectadas no tienen vías accesibles de reclamación. La gobernanza debe contemplar competencias, coordinación, formación, auditoría y mecanismos de reparación.

Para evaluar la calidad de un marco jurídico conviene considerar, al menos, si identifica con claridad a los responsables; si exige evaluaciones proporcionales al riesgo; si contempla documentación y trazabilidad; si protege los derechos de las personas afectadas; si prevé supervisión humana significativa; si facilita auditorías independientes; si ofrece vías de impugnación y reparación; y si establece coordinación entre autoridades. Estos criterios no constituyen una fórmula automática para resolver todos los casos. Sirven como guía para determinar

si la regulación conecta las capacidades técnicas con obligaciones verificables.

En América Latina, estas preguntas deben abordarse desde los marcos constitucionales y legales de cada país, además de los compromisos internacionales aplicables. La diversidad institucional, económica y lingüística de la región impide trasladar sin adaptación modelos extranjeros. Las políticas regulatorias deben considerar las capacidades de supervisión, la dependencia de proveedores externos, la disponibilidad de datos representativos y la participación de comunidades históricamente excluidas. La seguridad jurídica requiere proteger derechos y, a la vez, evitar que obligaciones impracticables consoliden la concentración tecnológica en pocos actores.

1.4. Enfoques metodológicos y teóricos para el estudio jurídico de la IA

1.4.1. Métodos de análisis jurídico y tecnológico

El estudio jurídico de la IA requiere combinar métodos sin confundir sus funciones. El método dogmático-jurídico permite identificar, interpretar y sistematizar normas, principios, competencias y categorías jurídicas vigentes. En una investigación sobre decisiones automatizadas, este método puede examinar las reglas constitucionales, administrativas, de protección de datos y de debido proceso aplicables en una jurisdicción concreta. Su aporte consiste en explicar qué dispone el derecho, cómo se relacionan sus fuentes y qué problemas de interpretación permanecen abiertos.

El método comparado permite estudiar cómo distintos ordenamientos regulan sistemas, riesgos y responsabilidades semejantes. La comparación rigurosa no consiste en enumerar leyes de varios países, sino en definir una pregunta común, justificar la selección de jurisdicciones y examinar sus contextos institucionales. Por ejemplo, puede compararse la forma en que determinados marcos establecen

evaluaciones de impacto o supervisión humana. El análisis debe distinguir normas vigentes, proyectos legislativos y políticas públicas, así como atender diferencias en jerarquía, alcance territorial y mecanismos de cumplimiento.

La investigación socio-jurídica resulta necesaria cuando se busca conocer cómo se implementan las normas en la práctica. Entrevistas a funcionarios, análisis de procedimientos, revisión de decisiones administrativas, estudios de caso y observación institucional pueden revelar diferencias entre las obligaciones formales y la experiencia de las personas afectadas. El enfoque empírico no reemplaza la interpretación normativa; permite observar cómo se aplica una regla, qué obstáculos enfrenta y qué consecuencias produce. Los diseños cualitativos son útiles para reconstruir procesos y significados, mientras que los cuantitativos permiten examinar patrones o diferencias en resultados, siempre que los datos y métodos sean adecuados.

El enfoque mixto puede integrar ambos tipos de evidencia. Una investigación podría analizar la normativa aplicable, revisar expedientes, entrevistar a personas usuarias y medir tasas de error o disparidad entre grupos. Para que la combinación sea válida, deben explicitarse la pregunta, las fuentes, los criterios de selección y las limitaciones. Las métricas de desempeño no deben interpretarse sin contexto: una tasa global de precisión puede ocultar errores concentrados en una población específica, y una asociación estadística no demuestra por sí sola causalidad ni discriminación jurídica.

El estudio técnico funcional es un complemento importante para el análisis legal. No exige que el investigador jurídico se convierta en desarrollador, pero sí que comprenda qué datos alimentan el sistema, qué objetivo se optimiza, qué resultados produce, qué variables intervienen y cómo se integra la salida en el procedimiento. La colaboración interdisciplinaria con especialistas en informática, estadística, ciencias sociales y ética puede reducir errores de caracterización. Debe

mantenerse, sin embargo, una distinción entre la explicación técnica de cómo funciona un modelo y la valoración jurídica de si su uso está permitido.

También es importante delimitar el objeto de estudio. "La IA en la justicia" puede referirse a un buscador de jurisprudencia, un sistema de transcripción, una herramienta de clasificación de expedientes o un modelo que recomienda medidas procesales. Cada sistema plantea cuestiones distintas. Una investigación sólida identifica el tipo de herramienta, la función institucional, los actores implicados, la población afectada, el periodo de análisis y la jurisdicción. Sin esa delimitación, las conclusiones tienden a ser generales y difíciles de verificar.

1.4.2. Marcos teóricos y preguntas de investigación

La teoría jurídica ayuda a explicar qué valores y categorías deben orientar la regulación. Los enfoques de derechos fundamentales examinan cómo la tecnología afecta dignidad, autonomía, igualdad, privacidad, libertad de expresión y tutela judicial efectiva. El análisis no debe quedarse en una lista de derechos: debe especificar qué interferencia se produce, qué autoridad o actor la causa, si existe una finalidad legítima, qué medidas de protección se aplican y qué vías de reparación están disponibles.

La teoría del Estado de derecho dirige la atención hacia la legalidad, la limitación del poder, la motivación de las decisiones y el control judicial. Desde esta perspectiva, la automatización plantea una cuestión institucional: ¿puede una autoridad ejercer válidamente una competencia si no comprende ni puede explicar los factores decisivos del sistema que emplea? La respuesta dependerá del tipo de competencia y de las reglas aplicables, pero el problema exige que la decisión permanezca atribuible a una autoridad legalmente habilitada y sujeta a control.

Los enfoques de justicia distributiva y no discriminación permiten analizar cómo se distribuyen beneficios y cargas. Un sistema puede reproducir desigualdades existentes si aprende de datos que reflejan prácticas históricas o si utiliza variables indirectas que correlacionan con características protegidas. También puede mejorar la consistencia en ciertos contextos. La evaluación debe precisar qué grupos se comparan, qué criterio de equidad se usa y qué consecuencias se consideran relevantes. No existe una única métrica de justicia que resuelva todas las controversias: la selección del criterio expresa decisiones normativas y debe justificarse.

Los estudios críticos de la tecnología cuestionan la idea de que las herramientas sean neutrales o autónomas respecto de las estructuras sociales. La selección de datos, la definición de objetivos y el diseño de interfaces reflejan prioridades institucionales y comerciales. Esta perspectiva dirige la atención hacia el poder de clasificar, medir y predecir, así como hacia la concentración de recursos y la dependencia de proveedores. Su aporte no es asumir que todo sistema es perjudicial, sino examinar quién controla la infraestructura, qué intereses orientan su uso y qué personas pueden cuestionar sus efectos.

El enfoque sociotécnico entiende que los resultados dependen de la interacción entre modelo, datos, personas, organizaciones, normas y contextos. Una herramienta con buen rendimiento técnico puede fallar al utilizarse en condiciones diferentes de las previstas; una decisión formalmente supervisada puede transformarse en aceptación automática; y un sistema que funciona adecuadamente en una institución puede producir efectos distintos en otra población. Por ello, las evaluaciones deben abarcar el ciclo de vida y las condiciones reales de despliegue, no solo el modelo aislado.

A partir de estos marcos pueden formularse preguntas de investigación más precisas. ¿Qué norma autoriza o limita el uso de un sistema en una determinada función? ¿Qué actores intervienen en el proceso y qué

deberes les corresponden? ¿La persona afectada recibe información comprensible y puede impugnar el resultado? ¿Qué datos y criterios intervienen en la clasificación? ¿Se observan impactos desiguales entre grupos? ¿Qué evidencia técnica y documental permite reconstruir la decisión? ¿Qué tipo de supervisión humana existe en la práctica? Las respuestas requieren conectar la dimensión normativa con la evidencia disponible.

La investigación también debe declarar sus límites. Si el acceso a datos o al código está restringido, debe explicarse qué inferencias son posibles y cuáles no. Si se usan métricas proporcionadas por el proveedor, conviene distinguirlas de una evaluación independiente. Si la regulación se encuentra en evolución, debe indicarse la fecha de corte y el estado de cada fuente. Este cuidado es especialmente importante en el derecho comparado, donde una norma aprobada puede no estar vigente, y en el estudio de tecnologías que cambian mediante actualizaciones frecuentes.

Un marco metodológico consistente evita dos reduccionismos. El primero considera que la IA es solo una cuestión técnica y deja fuera las relaciones de poder, derechos y responsabilidades. El segundo, trata el sistema como una abstracción jurídica sin examinar su funcionamiento. El análisis riguroso necesita describir técnicamente lo indispensable, interpretar el derecho aplicable y observar el contexto en que se produce el impacto. Así se puede valorar si una solución normativa responde al problema real y si existen condiciones institucionales para cumplirla.



Capítulo 2.

Fundamentos éticos y principios de la inteligencia artificial



INTRODUCCIÓN

La reflexión ética sobre la inteligencia artificial estudia cómo orientar su diseño, desarrollo, adquisición y uso para que sus beneficios sean compatibles con la dignidad humana, los derechos y el bienestar colectivo. No se limita a preguntar si una tecnología funciona: examina qué fines persigue, quién participa en su configuración, qué personas pueden resultar afectadas y cómo se distribuyen sus beneficios y riesgos. La ética tampoco sustituye al derecho. Puede anticipar problemas, orientar decisiones responsables y contribuir a la formulación de normas, pero no convierte por sí sola una recomendación en una obligación jurídica exigible.

Este capítulo analiza cuatro cuestiones conectadas. Primero presenta los principios éticos internacionales y sus fundamentos teóricos. Luego examina transparencia, explicabilidad y rendición de cuentas como condiciones para comprender y controlar los sistemas. En tercer lugar aborda sesgos, equidad y no discriminación. Finalmente estudia cómo los principios éticos pueden pasar a normas jurídicas vinculantes, así como el papel y los límites de la autorregulación y del *soft law*. El propósito es conservar una distinción clara entre valores orientadores, estándares voluntarios y obligaciones jurídicas aplicables en una jurisdicción determinada.

2.1. Ética de la IA: principios internacionales y corrientes teóricas

2.1.1. Fundamentos y convergencia de los principios

La ética de la IA no constituye una teoría única ni un código universal con respuestas automáticas. Es un campo interdisciplinario en el que confluyen filosofía moral, derecho, ciencias de la computación, ciencias sociales y políticas públicas. Su objeto comprende decisiones humanas relacionadas con sistemas computacionales: qué objetivos optimizar, qué datos emplear, qué usos permitir, qué riesgos aceptar y cómo responder cuando se producen consecuencias adversas. El análisis ético se ocupa, por tanto, tanto de las propiedades técnicas como de las condiciones sociales e institucionales en que se usan los sistemas.

En los últimos años, organismos internacionales, gobiernos, universidades y empresas han publicado marcos de principios. Jobin, Ienca y Vayena (2019) identificaron, en su revisión de directrices globales, una convergencia en torno a temas como transparencia, justicia, no maleficencia, responsabilidad y privacidad, aunque con diferencias en su interpretación y prioridad. La coincidencia terminológica no significa que exista acuerdo sobre su aplicación. "Justicia", por ejemplo, puede referirse a igual trato, reparación de desventajas históricas, distribución equitativa de beneficios o ausencia de resultados discriminatorios. Cada sentido lleva a criterios de evaluación diferentes.

La Recomendación de la UNESCO sobre la Ética de la Inteligencia Artificial, adoptada en 2021, coloca la dignidad y los derechos humanos en el centro de su marco. Incluye principios como proporcionalidad y prevención del daño, seguridad, privacidad y protección de datos, supervisión humana, transparencia, responsabilidad, equidad, inclusión y sostenibilidad. Su orientación comprende todo el ciclo de vida de los sistemas y propone medidas de política pública para que los Estados

traduzcan sus principios en prácticas. Se trata de una recomendación internacional y no de un tratado que imponga por sí solo obligaciones jurídicas directamente exigibles a todas las personas u organizaciones.

La OCDE también ofrece un marco de principios para una administración responsable de sistemas de IA confiables. Sus recomendaciones se dirigen a gobiernos y actores de la cadena de valor, y vinculan el desarrollo y uso de la IA con crecimiento inclusivo, bienestar, valores centrados en las personas, transparencia, robustez, seguridad y rendición de cuentas. Los principios fueron adoptados en 2019 y actualizados en 2024 para reflejar cambios tecnológicos y de política pública, incluidos los relacionados con la IA generativa. Como recomendación del Consejo de la OCDE, su fuerza principal es orientadora y política; su incorporación a una norma nacional puede producir efectos jurídicos distintos según el instrumento que la implemente.

Entre las propuestas teóricas más conocidas se encuentra el marco de Floridi y Cowls (2019), que sintetiza cinco principios: beneficencia, no maleficencia, autonomía, justicia y explicabilidad. El modelo conecta la aspiración de producir beneficios con el deber de evitar daños, preservar capacidad de decisión, distribuir de manera justa las consecuencias y hacer comprensible la actuación del sistema. Otros enfoques, como AI4People, desarrollan una arquitectura ética que incluye principios y recomendaciones para orientar a diseñadores, organizaciones y responsables de políticas (Floridi et al., 2018). Estas propuestas ayudan a ordenar el debate, pero no reemplazan el análisis de contexto ni establecen por sí mismas una regla de decisión aplicable a todo caso.

La convergencia en los principios refleja una preocupación común: la capacidad tecnológica debe mantenerse vinculada a fines humanos y a obligaciones de cuidado. Sin embargo, formular valores abstractos es solo un punto de partida. “Respetar la autonomía” requiere definir

qué información recibe una persona, si puede rechazar el uso de una herramienta y si tiene una alternativa efectiva. “Promover la justicia” requiere precisar qué grupos pueden resultar perjudicados y qué evidencia se empleará para detectar diferencias. “Prevenir daños” exige identificar riesgos previsibles, establecer controles y determinar quién tiene autoridad para detener o modificar el sistema.

2.1.2. Principios en tensión y aplicación contextual

Los principios éticos pueden entrar en tensión. Por ejemplo, mayor recopilación de datos puede mejorar el rendimiento de un modelo, pero afectar la privacidad. Una explicación amplia de una decisión puede facilitar la impugnación, pero exponer información protegida o permitir que actores maliciosos manipulen el sistema. La personalización puede mejorar un servicio para una persona, pero producir segmentación injusta o limitar su acceso a alternativas. La ética aplicada requiere reconocer estos conflictos y explicar por qué una solución prioriza ciertos valores, bajo qué condiciones y con qué salvaguardas.

El principio de beneficencia invita a orientar la IA hacia fines socialmente valiosos, como mejorar el acceso a servicios, ampliar capacidades o apoyar la investigación. No equivale a considerar beneficios comerciales o eficiencia organizacional como beneficios suficientes. La evaluación debe identificar quién recibe las ventajas, quién soporta los costos y si el sistema resuelve una necesidad real. Una herramienta puede mejorar el promedio de desempeño y, a la vez, dejar fuera a personas cuya lengua, discapacidad o contexto no fueron considerados en el diseño.

La no maleficencia requiere prevenir daños razonablemente previsibles y reducir sus efectos cuando no puedan eliminarse por completo. En IA, los daños pueden incluir decisiones erróneas, discriminación, pérdida de privacidad, vigilancia desproporcionada, desinformación o dependencia excesiva de recomendaciones automatizadas. El principio no exige suponer que cualquier riesgo vuelve inadmisible

una tecnología. Exige analizar la gravedad y probabilidad de los daños, la posibilidad de corregirlos y la existencia de alternativas menos intrusivas. La UNESCO articula esta idea mediante proporcionalidad y prevención del daño, vinculando el uso de sistemas con una finalidad legítima y con medidas adecuadas al riesgo.

La autonomía se relaciona con la capacidad de las personas para comprender, decidir y actuar sin manipulación ni coerción indebida. Un sistema puede afectar esa capacidad si emplea personalización opaca para influir en conductas, si presenta una recomendación como inevitable o si no ofrece vías razonables para rechazar una decisión automatizada. La supervisión humana puede contribuir a proteger la autonomía, pero su mera presencia formal no es suficiente. Quien supervisa debe contar con información, tiempo, competencia y autoridad para intervenir, y no limitarse a confirmar rutinariamente las salidas del sistema.

La justicia comprende tanto la distribución de oportunidades y cargas como la calidad del trato y la posibilidad de impugnar una decisión. La equidad no siempre equivale a aplicar la misma regla a todas las personas: un tratamiento idéntico puede perpetuar desventajas si los grupos enfrentan condiciones distintas. Pero las medidas diferenciadas también requieren justificación y consistencia con el marco jurídico. De ahí la necesidad de describir el problema, elegir criterios pertinentes y evaluar resultados en contexto, en vez de afirmar que un sistema es justo por cumplir una sola métrica.

La explicabilidad, en el plano ético, se conecta con la posibilidad de entender una decisión lo suficiente para valorar sus consecuencias y actuar frente a ella. No es una propiedad aislada del modelo, sino una relación entre la información ofrecida, quien la necesita y la decisión que debe comprenderse. Una persona afectada puede requerir razones comprensibles y una vía para corregir datos; un auditor puede necesitar documentación técnica; una autoridad puede requerir evidencia de

cumplimiento. Una explicación genérica, como “el resultado se generó mediante inteligencia artificial”, no satisface ninguna de estas necesidades.

Los principios tampoco deben convertirse en una lista de verificación que permita declarar “ética” a una organización por completar formularios. Mittelstadt et al. (2016) advierten que el debate sobre algoritmos incluye problemas diferentes, desde sesgos y privacidad hasta opacidad y efectos en la autonomía. Para aplicar los principios hay que asignar responsabilidades, identificar personas afectadas, prever mecanismos de monitoreo y crear vías de reclamación. La evaluación ética gana contenido cuando se conecta con decisiones documentadas y resultados observables.

La orientación centrada en las personas tampoco implica ignorar los efectos colectivos y ambientales. Los sistemas pueden alterar condiciones laborales, acceso a información, distribución de recursos y consumo de energía o agua. Un marco ético completo pregunta por impactos sobre comunidades, instituciones y generaciones futuras, además de las consecuencias para cada usuario. El principio de sostenibilidad incorporado en la Recomendación de la UNESCO amplía la evaluación hacia efectos sociales y ambientales del ciclo de vida.

Por último, los principios deben interpretarse junto con el derecho vigente. En algunos contextos, una obligación jurídica concreta puede ofrecer un umbral mínimo de protección; en otros, el deber ético puede motivar una salvaguarda adicional. Una práctica legal no es necesariamente ética, y una práctica considerada éticamente deseable no es automáticamente exigible en tribunales. Mantener esa distinción permite analizar con precisión qué se recomienda, qué se debe hacer y qué consecuencias puede generar el incumplimiento.

Tabla 2.1

Principios de IA, riesgo asociado y medida de aplicación

Principio	Riesgo asociado	Medida de aplicación
Dignidad y autonomía	Que las personas sean tratadas como perfiles o puntuaciones, sin capacidad real de decisión	Informar cuándo interviene la IA y permitir que la persona cuestione o revise el resultado
Beneficencia	Que el sistema no aporte el beneficio esperado o que este solo favorezca a determinados actores	Definir la finalidad y evaluar si el sistema mejora efectivamente el servicio o proceso
No maleficencia	Daños previsibles a derechos, seguridad, privacidad o bienestar	Identificar riesgos antes del uso, establecer medidas de mitigación y suspender el sistema si el daño no puede controlarse
Justicia y equidad	Errores o impactos desiguales entre grupos	Revisar los datos y resultados, analizar diferencias relevantes y corregir el sistema o el proceso
Privacidad y protección de datos	Recolección excesiva, uso incompatible o divulgación de información personal	Limitar los datos a los necesarios, protegerlos y respetar los derechos de sus titulares
Transparencia y explicabilidad	Que las personas desconozcan el uso del sistema o no comprendan los factores que influyeron en el resultado	Comunicar la finalidad, las limitaciones, los responsables y las vías de revisión con información adecuada para cada audiencia

Principio	Riesgo asociado	Medida de aplicación
Responsabilidad y rendición de cuentas	Que los actores atribuyan la responsabilidad a otros integrantes de la cadena	Asignar responsables, documentar decisiones y conservar registros proporcionales al riesgo
Seguridad y sostenibilidad	Fallas, usos indebidos o impactos sociales y ambientales no evaluados	Probar y monitorear el sistema durante su ciclo de vida, y revisar sus efectos sobre las personas y el entorno

Nota. *Los principios pueden entrar en tensión y su aplicación depende de la finalidad, el contexto y las personas afectadas. Elaboración propia con base en UNESCO (2021), Organisation for Economic Co-operation and Development (2024) y Floridi y Cowls (2019).*

2.2. Transparencia, explicabilidad y rendición de cuentas

2.2.1. Transparencia y explicabilidad: alcances y límites

La transparencia se refiere a la disponibilidad de información relevante sobre un sistema y su uso. Puede incluir la finalidad, los actores responsables, los datos empleados, las limitaciones conocidas, los procedimientos de supervisión y las condiciones bajo las cuales sus resultados influyen en decisiones. No exige publicar indiscriminadamente todo el código, los datos personales o los secretos comerciales. Su propósito es permitir que personas, organizaciones y autoridades comprendan lo necesario para evaluar el sistema, detectar fallos y ejercer control.

La información requerida depende del destinatario. Una persona afectada necesita saber que una herramienta influyó en la decisión que la concierne, qué factores principales fueron relevantes y cómo solicitar revisión. Una organización usuaria necesita instrucciones sobre el alcance, las limitaciones y el uso previsto del sistema. Un auditor puede requerir documentación técnica y datos de prueba.

Una autoridad supervisora puede necesitar registros completos para verificar cumplimiento. Presentar el mismo documento técnico a todos no asegura transparencia: la información debe ser pertinente, accesible y adecuada a las capacidades y necesidades de cada audiencia.

La explicabilidad está relacionada, pero no es idéntica. Describe la posibilidad de ofrecer razones comprensibles acerca de cómo una entrada, un modelo o determinados factores contribuyeron a un resultado. En algunos casos, el sistema es interpretable porque su estructura permite inspeccionar directamente la relación entre variables y salida. En otros, se utilizan métodos posteriores para aproximar o resumir la contribución de ciertos elementos. Una explicación aproximada puede ser útil, pero no debe confundirse con una reconstrucción perfecta del funcionamiento interno. Además, una explicación técnicamente plausible no demuestra que la decisión sea legal o justa.

Conviene diferenciar explicación global y explicación individual. La primera caracteriza tendencias generales del modelo: qué patrones suele detectar, qué variables utiliza o en qué condiciones falla. La segunda se refiere a un resultado concreto: por qué una solicitud fue denegada o qué elementos influyeron en la clasificación de una persona. Ambas cumplen fines distintos. Una descripción general del modelo puede servir para supervisarlo, mientras que una explicación individual puede permitir corregir datos, cuestionar factores improcedentes o solicitar una revisión humana.

La transparencia también abarca el uso institucional del sistema. La organización debe aclarar si el modelo es experimental o está en producción; quién puede modificar sus parámetros; qué control humano existe; qué incidencias se registran; y qué canales se ofrecen para reclamar. Sin esa dimensión organizativa, la publicación de información sobre el modelo puede dejar intacta la opacidad del proceso decisorio. Es posible conocer cómo se calculó una puntuación

y seguir sin saber quién decidió convertirla en una medida adversa.

La transparencia enfrenta límites legítimos. La protección de datos personales, la ciberseguridad y la propiedad intelectual pueden restringir la divulgación pública de ciertos detalles. Sin embargo, tales límites deben aplicarse de forma proporcionada y no utilizarse como una excusa genérica para impedir revisión. Puede ser posible compartir información con autoridades o expertos independientes bajo condiciones de confidencialidad; explicar a una persona afectada los factores relevantes sin revelar el código completo; o entregar documentación agregada que no exponga datos personales. La cuestión no es optar por divulgación total o secreto absoluto, sino definir mecanismos de acceso adecuados a cada finalidad.

En el contexto normativo europeo, el Reglamento (UE) 2024/1689 establece obligaciones de transparencia e información para determinados sistemas y usos, incluyendo requerimientos de información para quienes los despliegan y reglas para ciertos contenidos generados o manipulados por IA. También regula la supervisión humana en determinadas categorías de sistemas de alto riesgo. Estas obligaciones son parte de un régimen jurídico regional y deben leerse conforme a su ámbito de aplicación, sujetos y fechas pertinentes; no representan por sí solas un deber general idéntico para todas las jurisdicciones.

2.2.2. Rendición de cuentas, trazabilidad y supervisión

La rendición de cuentas significa que los actores deben poder explicar y justificar sus decisiones, asumir responsabilidades y responder ante mecanismos de supervisión o reparación. En la gobernanza de IA, no basta con demostrar que una herramienta fue adquirida de un proveedor o que el sistema produjo una salida. La institución debe poder explicar por qué adoptó la herramienta, qué controles implementó, cómo verificó su idoneidad y qué medidas tomó ante errores o impactos adversos.

La rendición de cuentas depende de la asignación de roles. Quien diseña el modelo puede controlar su arquitectura; quien lo ofrece en el mercado define documentación y soporte; quien lo integra puede adaptarlo a un proceso; y quien lo usa decide cómo interpretar sus salidas. Un actor puede acumular varios roles. La obligación concreta dependerá de la normativa aplicable y de su capacidad real de prevenir o corregir el riesgo. Una organización no debería eludir la responsabilidad que le corresponde alegando que el proveedor considera confidencial la información necesaria para supervisar el uso.

La trazabilidad es la capacidad de reconstruir decisiones y acciones a través de registros adecuados. Puede abarcar versiones del modelo, cambios de configuración, fuentes y transformaciones de datos, instrucciones de entrada, resultados, intervención humana y revisiones posteriores. No es necesario conservar indiscriminadamente todos los datos sin límite temporal; la documentación debe guardar relación con los riesgos, las obligaciones legales y las necesidades de auditoría. La ausencia de registros relevantes, sin embargo, puede impedir establecer qué ocurrió y dificultar la corrección de errores.

La rendición de cuentas también exige distinguir la supervisión nominal de la supervisión efectiva. Una persona no ejerce control significativo si solo recibe una puntuación final, no conoce sus límites y no puede apartarse de ella. El diseño organizativo debe permitir que el supervisor entienda el sistema en la medida necesaria para su función, identifique señales de fallo y detenga o escale decisiones cuando corresponda. La formación, el tiempo disponible y la independencia frente a presiones de productividad son condiciones prácticas de control.

La auditoría puede ser interna, externa o regulatoria. La auditoría interna ayuda a verificar procedimientos, detectar desviaciones y documentar decisiones; la externa añade independencia y puede aumentar la confianza, aunque depende de que exista acceso suficiente a la evidencia; y la supervisión pública permite evaluar cumplimiento

y aplicar medidas correctivas o sancionatorias conforme al derecho. Ningún método asegura por sí solo ausencia de daño. Las auditorías deben revisar tanto el rendimiento técnico como el contexto de despliegue, los datos, los impactos diferenciados y la eficacia de los canales de revisión.

El marco de gestión de riesgos de IA del Instituto Nacional de Estándares y Tecnología de Estados Unidos (NIST AI RMF 1.0) es una referencia voluntaria que propone integrar consideraciones de confiabilidad en el diseño, desarrollo, uso y evaluación de sistemas. Incluye características como validez y fiabilidad, seguridad, resiliencia, responsabilidad, transparencia, explicabilidad, protección de la privacidad y equidad con gestión de sesgos dañinos. NIST aclara que el marco está pensado para uso voluntario; su adopción por una entidad puede orientar prácticas, pero no lo convierte automáticamente en ley.

La transparencia y la rendición de cuentas se refuerzan entre sí. Sin información, no hay revisión significativa; sin una autoridad o actor responsable, la información no conduce a correcciones. Pero no son intercambiables: un sistema puede ser transparente y aun así no tener un responsable claramente identificado; también puede existir una entidad formalmente responsable que no pueda reconstruir el funcionamiento del modelo. Una gobernanza adecuada vincula información, registros, competencias, vías de impugnación y medidas de reparación.

2.3. Sesgos algorítmicos, equidad y no discriminación

2.3.1. Orígenes y formas de sesgo

El término “sesgo algorítmico” se utiliza para describir patrones sistemáticos de error o trato desigual asociados con sistemas computacionales. No todo error constituye discriminación jurídica, y no toda diferencia estadística revela por sí misma un sesgo injusto. El análisis requiere identificar qué resultado se considera adverso, qué grupo se ve afectado, cuál es el comparador pertinente y qué relación

existe entre la diferencia observada y los datos, el diseño o el uso del sistema.

Los sesgos pueden surgir en diversas etapas. En la selección de datos, una población puede estar infrarrepresentada; en la medición, la variable utilizada puede ser un sustituto inadecuado del concepto que se pretende evaluar; en el etiquetado, las categorías de referencia pueden incorporar juicios previos; en el diseño, el objetivo de optimización puede privilegiar eficiencia sobre equidad; y en el despliegue, el sistema puede utilizarse en un contexto distinto de aquel para el que fue validado. La interacción entre estas etapas hace insuficiente atribuir todos los problemas a "datos sesgados".

Los datos históricos pueden reflejar desigualdades previas. Si un modelo aprende de decisiones pasadas de contratación, crédito o vigilancia, puede reproducir patrones que fueron resultado de exclusión o de acceso desigual. El hecho de que el modelo imite regularidades observadas no justifica convertirlas en criterios futuros. La correlación puede ser predictiva y, al mismo tiempo, normativamente improcedente. Un sistema que estima resultados con mayor precisión promedio podría aún perjudicar a una minoría si comete errores más frecuentes en ese grupo o si la decisión afecta de modo más severo sus derechos.

Los sesgos también pueden aparecer por variables proxy. Aunque se excluya una característica protegida, otros datos —como ubicación, historial de consumo o trayectoria laboral— pueden correlacionarse con ella. Eliminar directamente una variable no garantiza neutralidad. A la inversa, recopilar información sensible para medir disparidades puede crear riesgos de privacidad y requerir una base jurídica apropiada. La evaluación debe equilibrar la detección de impactos con la protección de datos y aplicar controles al acceso, uso y conservación de esa información.

La automatización puede amplificar consecuencias porque aplica un

criterio de manera rápida y extensa. Un proceso manual también puede contener sesgos, pero la escala del sistema puede producir efectos uniformes sobre miles de casos, con apariencia de objetividad y menor visibilidad pública. La uniformidad no equivale a imparcialidad. Si el modelo opera con datos defectuosos o un criterio inadecuado, la consistencia puede multiplicar el daño en lugar de corregirlo.

Las fuentes de sesgo no se limitan a errores estadísticos. La representación social de ciertos grupos, las prioridades comerciales, la ausencia de participación de comunidades afectadas y la falta de diversidad en equipos de diseño pueden influir en la formulación del problema y en las opciones disponibles. La equidad requiere preguntar quién definió el objetivo, qué experiencias se consideraron y qué grupos pudieron advertir problemas antes de la implementación. La consulta no elimina automáticamente el sesgo, pero puede revelar impactos que no son visibles desde la perspectiva del proveedor o la entidad usuaria.

2.3.2. Equidad, no discriminación y gestión del riesgo

La equidad es un principio ético amplio; la no discriminación es también una categoría jurídica cuyo contenido depende de las normas aplicables. En términos generales, la evaluación de una posible discriminación considera la existencia de una distinción, exclusión o restricción relacionada con una característica protegida o con un criterio que produce un impacto adverso. La legalidad del trato puede depender de su finalidad, justificación, proporcionalidad y efectos, según el régimen pertinente. Por ello, un análisis técnico de diferencias entre grupos es una fuente de evidencia, pero no reemplaza la calificación jurídica.

Las métricas de equidad pueden ofrecer información sobre el comportamiento del sistema, aunque no existe una métrica que resuelva todas las situaciones. La igualdad de tasas de error, la igualdad de tasas de selección y la igualdad de valores predictivos pueden entrar en conflicto cuando las distribuciones de datos y las tasas de referencia

difieren. La elección de una medida implica una decisión sobre qué tipo de error se considera más importante y qué población debe protegerse. La selección debe justificarse según la finalidad, el daño potencial y los principios jurídicos pertinentes.

La evaluación también debe considerar el impacto individual, no solo el promedio del grupo. Un resultado estadístico aparentemente equilibrado puede ocultar casos particulares con consecuencias graves. Asimismo, la categoría de grupo puede no representar identidades múltiples o interseccionales; una evaluación por género y otra por edad, por ejemplo, podrían no detectar impactos específicos sobre mujeres mayores o personas no binarias. El análisis desagregado es valioso, pero debe evitar conclusiones basadas en muestras pequeñas o información incompleta.

La gestión ética del sesgo comienza antes de desplegar el sistema. Requiere evaluar si la automatización es necesaria y adecuada para el objetivo; revisar la calidad, pertinencia y representatividad de los datos; documentar decisiones de diseño; probar resultados en condiciones realistas; e involucrar a especialistas y partes afectadas. Durante el uso, se deben vigilar cambios de población, contexto o rendimiento, registrar incidentes y activar revisiones cuando aparecen disparidades relevantes. La evaluación posterior debe poder conducir a cambios en el modelo, el proceso o la decisión de no continuar con el sistema.

La supervisión humana puede ayudar a detectar errores y considerar circunstancias que el modelo no representa. Sin embargo, no debe usarse como respuesta automática a todo riesgo. Si la persona supervisora reproduce la recomendación sin análisis o no puede cuestionarla, la intervención no mejora la equidad. Además, la persona afectada puede necesitar información, revisión independiente y una oportunidad real para presentar pruebas o corregir datos. La responsabilidad institucional abarca tanto la salida del sistema como el procedimiento en el que se utiliza.

La equidad no se logra siempre mediante la misma solución técnica. Puede requerir mejorar datos, cambiar variables, redefinir el objetivo, fijar límites al uso, ofrecer alternativas no automatizadas o retirar la herramienta. También puede exigir medidas de accesibilidad, asistencia lingüística y participación de comunidades. La decisión debe considerar la gravedad de los derechos comprometidos y las alternativas disponibles; no es suficiente afirmar que el sistema supera una comparación con el desempeño humano si el procedimiento sigue siendo injusto o carece de revisión.

La Recomendación de la UNESCO incluye equidad y no discriminación entre sus principios, y la OCDE relaciona una IA confiable con respeto a derechos, valores democráticos y prevención de impactos injustos. Estos estándares orientan la evaluación, aunque la obligación jurídica específica de prevenir o reparar discriminación proviene de las normas vinculantes aplicables en cada jurisdicción. El análisis debe identificar esas fuentes nacionales y regionales sin presentar el principio ético como sustituto del derecho antidiscriminatorio.

2.4. De la ética al derecho: soft law, autorregulación y su fuerza normativa

2.4.1. Diferencias entre ética, derecho y soft law

Los principios éticos, las normas jurídicas y el *soft law* pueden tratar temas similares, pero difieren en fuente, fuerza y mecanismos de cumplimiento. Un principio ético orienta la valoración de conductas y puede influir en decisiones de diseño o política pública. Una norma jurídica, aprobada por una autoridad competente o reconocida como fuente del derecho, establece obligaciones, prohibiciones o facultades que pueden ser exigibles mediante procedimientos institucionales. El *soft law* comprende instrumentos no vinculantes en sentido estricto —como recomendaciones, declaraciones, marcos de gobernanza y códigos voluntarios— que pueden influir en prácticas, interpretación normativa y reformas futuras.

La fuerza jurídica de un instrumento no se determina solo por su título. Una guía técnica puede ser voluntaria, mientras que una disposición que incorpora su estándar en una ley, reglamento, contrato o condición de licencia puede producir efectos exigibles en ese contexto. Un documento de orientación emitido por una autoridad puede influir en la evaluación de diligencia o en la interpretación de obligaciones generales, sin convertirse por ello en una ley. El análisis debe identificar quién emitió el instrumento, a quién se dirige, qué mecanismos de seguimiento contiene y si ha sido incorporado por una fuente vinculante.

El *soft law* puede ser útil en un entorno de rápida evolución porque facilita acuerdos, aprendizaje y actualización más ágil que los procesos legislativos. Puede traducir principios generales en prácticas recomendadas, establecer vocabulario compartido y orientar organizaciones que todavía no están sujetas a reglas específicas. Su flexibilidad también puede estimular cooperación internacional, especialmente cuando los ordenamientos nacionales difieren. No obstante, la flexibilidad puede derivar en vaguedad, aplicación desigual y ausencia de reparación para quienes resulten perjudicados.

La Recomendación de la UNESCO, los Principios de IA de la OCDE y el NIST AI RMF ilustran instrumentos de distinta naturaleza dentro de la gobernanza ética y técnica. La recomendación de UNESCO ofrece un marco internacional sobre ética y políticas; los principios de la OCDE articulan una orientación intergubernamental; y el marco NIST proporciona una metodología voluntaria de gestión de riesgos. Ninguno debe confundirse con una ley directamente aplicable en todos los Estados. Su implementación puede ocurrir mediante estrategias nacionales, normas sectoriales, contratos, políticas institucionales o legislación. AI

El tránsito de un principio al derecho suele ser gradual. La transparencia, por ejemplo, puede aparecer primero en declaraciones éticas; después,

en guías que recomiendan documentar sistemas; y finalmente, en normas que exigen información concreta para determinados usos. Este tránsito no es automático ni siempre deseable en la misma forma. La legislación debe decidir qué sujetos quedan obligados, qué contenido se exige, qué excepciones son legítimas, qué autoridad supervisa y qué consecuencias siguen al incumplimiento. Sin esta especificación, la incorporación de un principio en una norma puede conservar una ambigüedad similar a la del marco voluntario.

Los instrumentos internacionales tampoco tienen idéntica fuerza. Una recomendación de una organización internacional orienta a los Estados, pero no equivale a un tratado ratificado. Un tratado crea obligaciones internacionales para sus partes conforme a sus disposiciones y a las reglas de vigencia. Una norma regional puede ser vinculante dentro de su ámbito territorial y material, como sucede con el Reglamento de IA de la Unión Europea para los sujetos comprendidos en su alcance. La comparación jurídica debe describir con precisión estos efectos.

2.4.2. Autorregulación, co-regulación y garantías de cumplimiento

La autorregulación consiste en que organizaciones o sectores establezcan políticas, principios, códigos y controles propios. Puede adaptarse a procesos específicos y promover aprendizaje interno, pero sus incentivos pueden ser insuficientes cuando la entidad debe evaluarse a sí misma o cuando los costos de prevención recaen en terceros. Un código empresarial de IA responsable no garantiza que haya auditoría independiente, reparación efectiva o sanción por incumplimiento. La credibilidad depende de la precisión de sus compromisos, la transparencia de resultados y la existencia de consecuencias verificables.

La co-regulación combina normas públicas con instrumentos desarrollados o aplicados por actores privados y técnicos. El legislador puede fijar objetivos y obligaciones generales, mientras que estándares,

guías y evaluaciones de conformidad concretan requisitos técnicos. Esta combinación puede permitir que los criterios técnicos evolucionen sin reformar continuamente la ley. Pero la autoridad pública debe conservar capacidad de supervisión y no delegar decisiones fundamentales sobre derechos a entidades sin legitimidad o rendición de cuentas suficientes.

La adopción de principios éticos en una organización necesita estructuras de gobernanza. Entre ellas pueden incluirse evaluaciones de impacto, comités con competencias claras, protocolos de revisión humana, canales para reportar incidentes, capacitación, auditorías y procedimientos de respuesta. La existencia de un comité no es suficiente si carece de acceso a información o autoridad para detener un despliegue. Las políticas internas deben traducirse en responsabilidades asignadas y procesos documentados a lo largo del ciclo de vida.

Las evaluaciones de impacto pueden ayudar a convertir principios generales en preguntas concretas: cuál es el objetivo legítimo del sistema; qué grupos podrían verse afectados; qué datos se usarán; qué errores son previsibles; qué alternativas menos intrusivas existen; qué medidas de supervisión se requieren; y cómo se atenderán reclamos. La evaluación debe ser proporcional al riesgo y repetirse cuando cambian la finalidad, la población, los datos o el modelo. Su función no es garantizar que nunca ocurra un daño, sino anticipar y reducir riesgos, justificar la decisión de uso y hacer posible la supervisión.

La autorregulación también plantea el riesgo de "ética de fachada", en la que las declaraciones públicas se usan para generar confianza sin cambios sustantivos en la práctica. Esto puede ocurrir si los compromisos no tienen indicadores, si los resultados negativos no se publican o si la evaluación se limita al cumplimiento documental. La respuesta requiere mecanismos externos y participativos, vías de denuncia protegidas, documentación verificable y control público. La participación de las personas afectadas debe realizarse antes y durante el despliegue, no solo después de que se produzca un perjuicio.

La transformación de principios en obligaciones jurídicas debe respetar el debido proceso legislativo y la claridad normativa. No todos los principios deben traducirse en reglas idénticas para cada sistema. Algunas obligaciones pueden ser generales —por ejemplo, mantener registros o asignar responsabilidades— y otras depender del riesgo, el sector o el tipo de decisión. El legislador debe especificar excepciones y mecanismos de revisión, evitar duplicaciones entre autoridades y asegurar que las exigencias puedan cumplirse y verificarse.

Finalmente, ética y derecho deben operar de manera complementaria, conservando sus diferencias. La ética permite examinar fines, consecuencias y responsabilidades que pueden no estar todavía resueltos por normas específicas. El derecho fija obligaciones mínimas, protege derechos y ofrece procedimientos de control y reparación. El *soft law* puede servir de puente entre ambos, siempre que no se presente como equivalente a una norma vinculante ni como sustituto de la supervisión pública. La calidad de la gobernanza depende de que los principios se traduzcan en decisiones responsables, reglas claras y remedios efectivos para las personas afectadas.

Figura 2.1.

De los principios éticos a las garantías verificables



Nota: *Elaboración propia. Debe distinguirse visualmente entre normas vinculantes, recomendaciones y medidas voluntarias.*



Capítulo 3.

Derechos fundamentales en la era algorítmica



INTRODUCCIÓN

La inteligencia artificial puede influir en el ejercicio de derechos fundamentales cuando participa en decisiones sobre personas, organiza el acceso a información o servicios, o interviene en espacios de trabajo, educación y participación pública. Esa influencia no convierte automáticamente a cada sistema en una amenaza ni hace que una decisión algorítmica sea jurídicamente inválida. Sí exige examinar quién lo utiliza, con qué finalidad, qué personas pueden verse afectadas, qué garantías existen y qué vías de revisión están disponibles.

Este capítulo analiza la dignidad, la autonomía y la libertad frente a los sistemas automatizados; la igualdad y la no discriminación; la libertad de expresión, el acceso a la información y el pluralismo; y, finalmente, los derechos laborales y sociales y la situación de grupos vulnerables. El análisis parte de una premisa: los derechos humanos siguen siendo aplicables en entornos mediados por tecnología. Por tanto, el uso de IA debe evaluarse conforme a las obligaciones jurídicas vigentes en cada jurisdicción y a los principios de necesidad, proporcionalidad, igualdad y tutela efectiva.

3.1. Dignidad humana, autonomía y libertad frente a los sistemas automatizados

3.1.1. Dignidad humana y reconocimiento de la persona

La dignidad humana constituye un fundamento de los derechos y un criterio para valorar el trato que reciben las personas en las decisiones institucionales. En el contexto de la IA, protege frente a prácticas que reducen a una persona a una puntuación, un perfil o una categoría sin atender a su situación individual. Esta consideración no implica prohibir el uso de estadísticas o herramientas de apoyo. Exige que el resultado computacional no sustituya la condición de sujeto de derechos ni convierta una estimación probabilística en una valoración definitiva sobre la identidad, el mérito o la conducta de alguien.

Un sistema de clasificación puede ser útil para organizar información, pero su uso resulta problemático cuando una categoría se trata como una característica permanente de la persona o como fundamento suficiente para limitar derechos. Una puntuación crediticia, una predicción de riesgo o una estimación de rendimiento no son hechos independientes de las decisiones que dieron forma al modelo. Dependen de datos, variables, objetivos y umbrales elegidos por personas u organizaciones. Si se presentan como verdades neutrales, se ocultan las decisiones que deben ser justificadas y se dificulta que las personas las cuestionen.

La dignidad también se relaciona con la privacidad, la identidad y el control sobre la propia representación. Los sistemas de perfilamiento pueden inferir hábitos, preferencias o características sensibles a partir de información que una persona no entregó directamente para ese propósito. Aunque una inferencia no sea correcta, puede producir consecuencias reales si se utiliza para negar un servicio, modular un precio o intensificar la vigilancia. El problema no reside únicamente en la recopilación de datos, sino también en las conclusiones que se extraen y en el poder que esas conclusiones adquieren dentro de una institución.

El enfoque de derechos humanos requiere examinar el ciclo completo de la tecnología. Una práctica respetuosa de la dignidad no puede evaluarse solo por la interfaz que ve el usuario o por la precisión promedio del modelo. También importan la forma en que se obtuvieron los datos, los fines del sistema, la población sobre la que se aplica, las alternativas existentes y la posibilidad de obtener una explicación y una reparación. La Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos ha advertido que tecnologías como la IA, el aprendizaje automático, el perfilamiento y la biometría pueden afectar la privacidad y otros derechos si se aplican sin salvaguardias adecuadas.

En el ámbito estatal, la exigencia es especialmente estricta porque las autoridades ejercen competencias públicas y deben actuar con base legal, finalidad legítima y respeto a las garantías aplicables. Una decisión administrativa no deja de ser imputable al Estado por haber sido apoyada por un modelo de un proveedor privado. La autoridad debe conservar la capacidad de conocer el procedimiento, explicar los criterios relevantes y responder ante las personas afectadas. En los usos privados, aunque la estructura de obligaciones varíe según la jurisdicción, también deben considerarse la dignidad, la igualdad y los deberes legales que rigen la actividad.

La protección de la dignidad no debe formularse como una resistencia abstracta a la tecnología. Un sistema puede ampliar capacidades, mejorar accesibilidad o facilitar el acceso a servicios. La evaluación debe identificar cuándo ese apoyo se convierte en sustitución indebida del juicio individual o en una forma de trato que impide a las personas participar en decisiones que las afectan. La persona debe seguir siendo reconocida como agente y titular de derechos, no únicamente como objeto de cálculo.

3.1.2. Autonomía, libertad y decisión automatizada

La autonomía comprende la posibilidad de formar y expresar decisiones propias, actuar con información suficiente y ejercer control sobre aspectos relevantes de la vida personal. Los sistemas automatizados pueden apoyar esa capacidad al facilitar información, traducción, asistencia y servicios personalizados. También pueden reducirla cuando manipulan la conducta, restringen opciones sin explicación o inducen a confiar excesivamente en recomendaciones que no son transparentes ni verificables.

La autonomía se ve comprometida cuando una persona no sabe que un sistema participa en una decisión que la afecta o no puede corregir información incorrecta. En determinadas actividades, una persona puede aceptar el uso de una herramienta porque no dispone de una alternativa real: por ejemplo, cuando el servicio esencial solo se ofrece mediante un canal digital o cuando el empleador utiliza un sistema automatizado como condición de acceso al puesto. La existencia formal de una opción no garantiza libertad efectiva si rechazarla supone perder un beneficio básico o una oportunidad relevante.

El consentimiento no es una solución universal a todos los problemas. Puede ser importante para el tratamiento de datos, pero no necesariamente legitima una práctica que afecta derechos de terceros, genera discriminación o se aplica en una relación de poder desigual. Además, aceptar términos extensos no significa comprender cómo un modelo inferirá características o influirá en decisiones. La protección de la autonomía requiere información comprensible, posibilidad de elección cuando resulte pertinente, límites a la manipulación y vías para impugnar consecuencias.

La intervención humana tampoco garantiza por sí sola una decisión autónoma. Si el funcionario o profesional recibe una recomendación como respuesta aparentemente objetiva, puede inclinarse a seguirla, incluso cuando carezca de fundamento suficiente. Esta dependencia

puede surgir de la carga laboral, la presión por cumplir indicadores o la falta de conocimientos técnicos. Por ello, una supervisión efectiva debe asegurar que la persona encargada pueda comprender las limitaciones, revisar los factores pertinentes y apartarse del resultado cuando el caso lo requiera.

La libertad incluye también la posibilidad de no ser sometido a vigilancia o perfilamiento injustificados. El uso de herramientas de seguimiento puede modificar la conducta, incluso cuando no se adopta una sanción directa. En el trabajo, la educación o los espacios públicos, la monitorización constante puede afectar la intimidad y generar autocensura. El examen jurídico debe considerar la base legal, la finalidad, el alcance, la duración y las salvaguardias, además de valorar si existen medios menos intrusivos para alcanzar el objetivo.

La autonomía también tiene una dimensión colectiva. Comunidades y grupos deben poder participar en las decisiones que afectan sus datos, territorios y formas de vida, particularmente cuando sistemas se despliegan sobre poblaciones con menor capacidad de influir en el diseño o cuestionar a las instituciones. La participación puede mejorar la identificación de riesgos y evitar que los criterios de quienes desarrollan o adquieren la tecnología se presenten como universales.

El marco de derechos humanos no exige que las personas comprendan cada parámetro matemático de un sistema. Exige que puedan conocer lo esencial sobre su participación en una decisión, acceder a una revisión adecuada y conservar la posibilidad de presentar información contextual. La responsabilidad de asegurar estas condiciones corresponde a los actores que diseñan, ofrecen y despliegan la herramienta, conforme a la legislación aplicable. Así, la autonomía se protege no solo mediante opciones individuales, sino también mediante procesos institucionales comprensibles y contestables.

3.2. Igualdad, no discriminación y justicia algorítmica

3.2.1. Igualdad ante la ley e impactos diferenciados

La igualdad ante la ley y la prohibición de discriminación exigen que las personas puedan ejercer sus derechos sin distinciones injustificadas. La tecnología puede contribuir a una aplicación más consistente de criterios, pero también puede reproducir desigualdades presentes en los datos, las instituciones o la sociedad. Un sistema automatizado no se vuelve imparcial por utilizar cálculos matemáticos: sus resultados dependen de cómo se definió el problema, qué variables se seleccionaron, qué datos se consideraron adecuados y qué errores se aceptaron.

La discriminación puede ser directa cuando una decisión emplea explícitamente una característica protegida para perjudicar a una persona o grupo, según el marco jurídico aplicable. También puede ser indirecta cuando un criterio aparentemente neutral produce una desventaja particular para un grupo, sin justificación suficiente. En los sistemas de IA, esta segunda forma puede aparecer, aunque la variable protegida se excluya del conjunto de datos, porque otras variables —como zona de residencia, historial educativo o patrones de consumo— pueden funcionar como sustitutos correlacionados.

La evaluación requiere ir más allá de la precisión global. Un modelo puede mostrar buen rendimiento promedio y cometer más errores en determinados grupos, o causar daños más graves cuando falla sobre algunas personas. Por ejemplo, un sistema de reconocimiento puede funcionar peor con ciertos tonos de piel; un filtro de selección laboral puede favorecer perfiles que se parecen a quienes fueron contratados anteriormente; una herramienta de evaluación educativa puede interpretar de manera incorrecta respuestas formuladas en una variedad lingüística poco representada. Los resultados deben analizarse por grupos pertinentes y en el contexto real de uso.

Los datos históricos presentan un desafío particular. Si reflejan decisiones institucionales previas marcadas por exclusión, el modelo puede convertir esas regularidades en criterios para decisiones futuras. La predicción de que un grupo recibió menos oportunidades no justifica mantener esa distribución. Debe preguntarse si los datos describen una realidad que debe corregirse o si representan un indicador legítimo para el objetivo. La correlación estadística, aun cuando sea sólida, no constituye por sí sola una justificación jurídica o moral.

La igualdad sustantiva también requiere atender las condiciones de acceso. Cuando el diseño presupone conexión estable, alfabetización digital, dominio de una lengua mayoritaria o disponibilidad de dispositivos compatibles, las personas que carecen de esas condiciones pueden quedar excluidas. Una interfaz automatizada puede ser formalmente igual para todos, pero producir obstáculos distintos para personas con discapacidad, comunidades rurales o hablantes de lenguas indígenas. La accesibilidad, los canales alternativos y la asistencia adecuada forman parte de la evaluación de igualdad.

La discriminación algorítmica puede surgir por una cadena de decisiones, no por una sola variable. En la preparación de datos pueden excluirse ciertos grupos; durante el entrenamiento, una función de optimización puede privilegiar el resultado mayoritario; y en el despliegue, una institución puede aplicar el sistema en un contexto diferente. Por eso, las medidas correctivas deben cubrir tanto el modelo como la organización que lo usa. Revisar el código sin examinar las prácticas institucionales puede dejar intacta la causa del impacto desigual.

3.2.2. Evaluación jurídica y salvaguardias

La evaluación de igualdad debe identificar la decisión, la población afectada, el criterio aplicado y la consecuencia. Después corresponde determinar qué normas nacionales, constitucionales o internacionales resultan aplicables y qué justificación puede ofrecer el actor responsable. La calificación jurídica depende de la jurisdicción y del

sector: las reglas pueden diferir respecto de empleo, educación, crédito, servicios públicos o acceso a bienes. Por ello, no conviene afirmar que cualquier diferencia estadística constituye automáticamente una infracción, ni que un resultado de alta precisión elimina la posibilidad de discriminación.

Las métricas de equidad pueden detectar diferencias, pero no resuelven por sí solas qué debe considerarse justo. Algunas evalúan tasas de error; otras, proporciones de selección o consistencia en predicciones. En ciertos contextos no es posible maximizar todas las métricas a la vez. Elegir cuál priorizar requiere atender la finalidad del sistema, la gravedad de los errores, la posición de los grupos y los derechos comprometidos. La decisión debe documentarse y explicarse, en lugar de presentarse como una consecuencia puramente técnica.

La evaluación previa al despliegue debería verificar si la automatización es necesaria y adecuada, si existen alternativas menos lesivas y si los datos representan razonablemente a las personas sobre las que se usará el sistema. Las pruebas deben emplear escenarios realistas, incluir grupos que podrían quedar infrarrepresentados y comprobar tanto falsos positivos como falsos negativos. También es necesario establecer umbrales para detener o revisar el sistema, comunicar sus límites y definir quién responderá ante una disparidad identificada.

La supervisión debe continuar después de la implementación. La población, las prácticas institucionales y los datos pueden cambiar; por tanto, un rendimiento aceptable en el momento de la adquisición no garantiza resultados adecuados en el tiempo. La organización debe monitorizar incidentes, comprobar diferencias entre grupos y revisar periódicamente la vigencia de las variables y objetivos. Cuando aparece un efecto adverso, puede ser necesario corregir datos, ajustar el sistema, modificar el proceso decisorio, ofrecer una revisión individual o suspender el uso.

Las garantías procesales son esenciales para que la igualdad sea exigible. Una persona afectada debe poder conocer que un sistema contribuyó a la decisión, corregir datos inexactos, presentar circunstancias relevantes y acceder a una revisión con capacidad real de modificar el resultado. En ciertos casos, puede ser necesario conservar evidencia técnica o permitir una auditoría independiente, respetando al mismo tiempo la privacidad y otros derechos. La falta de acceso a información no debe transformar la opacidad del sistema en una barrera absoluta para reclamar.

Las medidas de equidad deben tener en cuenta la intersección de distintas características. Una evaluación separada por género, edad o discapacidad puede no identificar daños que se concentran en personas que reúnen varias de esas condiciones. La desagregación estadística puede ayudar, pero requiere muestras suficientes y protección de datos sensibles. Cuando la evidencia cuantitativa es limitada, entrevistas, participación comunitaria y análisis cualitativo pueden revelar barreras que no aparecen en las métricas.

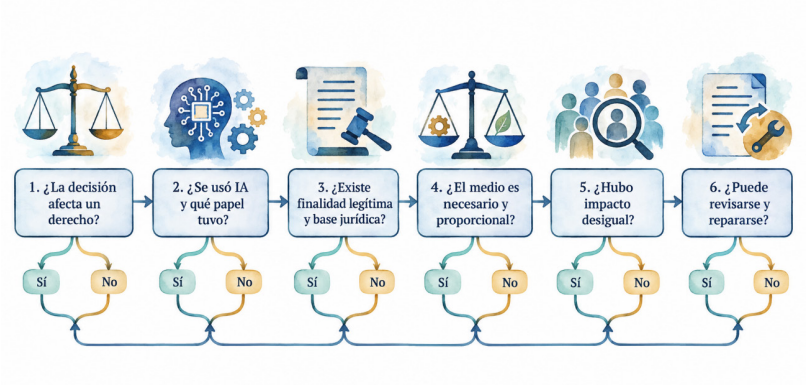
La igualdad también se relaciona con la distribución de beneficios. Si los sistemas de IA se entrenan con datos de comunidades que luego no reciben mejoras en servicios o acceso, pueden consolidarse relaciones asimétricas. La evaluación ética y jurídica debe considerar quién controla los datos, quién obtiene valor económico y quién soporta los riesgos. El enfoque de derechos humanos exige que la innovación no profundice exclusiones y que los grupos afectados participen en las decisiones que determinan el uso de la tecnología.

La respuesta contra la discriminación no se reduce a modificar el algoritmo. Puede requerir cambios en las reglas institucionales, los procesos de contratación, la formación del personal, la atención a usuarios o la distribución de recursos. A veces la medida más adecuada es retirar el sistema si su uso no puede justificarse o sus efectos no pueden corregirse. La tecnología debe permanecer subordinada a las

obligaciones de igualdad y no discriminación, y no convertirse en un motivo para diluirlas.

Figura 3.1.

Evaluación de una afectación de derechos por una decisión algorítmica



Nota: *Elaboración propia. Presentarlo como guía de análisis, no como sustituto de la evaluación jurídica del caso.*

3.3. Libertad de expresión, información y pluralismo en entornos algorítmicos

3.3.1. Expresión y acceso a información

La libertad de expresión comprende la posibilidad de buscar, recibir y difundir informaciones e ideas por distintos medios. En la actualidad, plataformas digitales y sistemas de recomendación organizan una parte importante de la circulación de contenidos. La clasificación automatizada puede ampliar el acceso a información, facilitar traducciones y ayudar a descubrir publicaciones. A la vez, puede influir en qué contenidos se muestran, cuáles pierden visibilidad y qué perspectivas alcanzan a diferentes públicos.

El artículo 19 del Pacto Internacional de Derechos Civiles y Políticos protege la libertad de opinión y expresión, sujeto a las restricciones previstas por el propio instrumento. La Observación general núm. 34 del Comité de Derechos Humanos explica la interpretación de ese derecho y aborda la libertad de expresión en medios contemporáneos. Al aplicarse a entornos mediados por IA, el análisis debe considerar si una restricción tiene base legal, persigue una finalidad legítima y es necesaria y proporcional. La automatización de la moderación no elimina estos requisitos ni convierte la decisión privada de una plataforma en una cuestión exenta de escrutinio jurídico cuando se ven afectados derechos.

Los sistemas de recomendación pueden generar personalización, pero también concentrar la exposición en contenidos que maximizan atención o interacción. Este diseño puede favorecer mensajes sensacionalistas, polarizantes o emocionalmente intensos, aunque el efecto depende de la plataforma, la configuración y el contexto. La evaluación de derechos debe examinar criterios de clasificación, opciones de control, mecanismos de apelación y el impacto sobre la diversidad de voces. No basta con afirmar que el contenido permanece técnicamente disponible si el sistema reduce sistemáticamente su alcance sin una explicación comprensible.

La moderación automatizada introduce otro conjunto de problemas. Los modelos pueden retirar contenido permitido, dejar visible contenido dañino o aplicar reglas de manera inconsistente entre idiomas y contextos culturales. Los falsos positivos pueden silenciar a periodistas, defensores de derechos humanos o comunidades que utilizan expresiones no reconocidas por el sistema; los falsos negativos pueden permitir amenazas o acoso. Las decisiones de moderación deben tener procedimientos de notificación, motivación y recurso, especialmente cuando afectan de forma significativa la participación pública o el acceso a información.

La escala de la moderación automatizada puede ser necesaria para gestionar grandes volúmenes, pero debe acompañarse de controles humanos y mecanismos de reparación. La revisión humana no debe limitarse a confirmar la clasificación inicial. Debe permitir considerar contexto, ironía, significado local y valor público del contenido. Además, los sistemas deben ser evaluados en las lenguas y comunidades en que se usan. La falta de datos lingüísticos adecuados puede producir una protección desigual: contenidos expresados en idiomas mayoritarios reciben análisis más preciso que los expresados en lenguas con pocos recursos digitales.

La IA generativa intensifica estas cuestiones porque puede producir textos, imágenes, voces y videos que aparentan ser auténticos. Puede facilitar expresión creativa y acceso a herramientas de comunicación, pero también permitir suplantaciones, falsificaciones o campañas de manipulación. Las respuestas deben proteger la libertad de expresión y, a la vez, atender daños concretos como fraude, acoso o interferencias en procesos públicos. Las restricciones generales basadas en la mera posibilidad de abuso pueden limitar de manera desproporcionada la creación legítima; la evaluación debe centrarse en usos, daños y salvaguardias verificables.

3.3.2. Pluralismo, manipulación y vigilancia

El pluralismo informativo exige que las personas puedan acceder a fuentes diversas y que el espacio público no quede controlado por un número reducido de actores o por sistemas de clasificación que invisibilizan determinadas perspectivas. La personalización puede facilitar que cada usuario encuentre información pertinente, pero también reducir encuentros con puntos de vista diferentes o reforzar preferencias existentes. No todo filtro produce necesariamente una "burbuja", y los efectos dependen de factores individuales y de plataforma; aun así, la arquitectura de recomendación merece escrutinio cuando influye de manera importante en la visibilidad pública.

La concentración de datos, infraestructura y capacidad computacional puede afectar la diversidad de expresión al otorgar a ciertos actores mayor poder para determinar las condiciones de circulación de información. Este problema combina libertad de expresión con competencia económica y gobernanza de plataformas. Las respuestas regulatorias deben ser precisas: transparencia sobre criterios, acceso a mecanismos de elección, auditoría independiente y protección frente a discriminación de contenidos pueden contribuir al pluralismo sin imponer una obligación de revelar información que facilite abusos o comprometa la seguridad.

La desinformación plantea un dilema complejo. La difusión de contenidos falsos puede causar daños y erosionar la confianza, pero atribuir al Estado o a una plataforma la facultad de decidir ampliamente qué es verdadero puede derivar en censura o control de opiniones. Las medidas deben distinguir entre expresión protegida, contenido ilegal y prácticas de manipulación coordinada, con criterios claros y vías de impugnación. Las herramientas de detección automatizada pueden servir como apoyo, pero no deberían operar como árbitro infalible de la verdad.

La vigilancia algorítmica puede afectar la libertad de expresión mediante monitoreo, perfilamiento o identificación de participantes en debates y protestas. Las personas pueden evitar expresarse si perciben que sus opiniones, contactos o desplazamientos quedan registrados y analizados. Este efecto inhibidor puede alcanzar también a periodistas y organizaciones de la sociedad civil. El derecho a expresar ideas debe protegerse junto con la privacidad y la libertad de reunión, de acuerdo con las garantías y restricciones que establezca el ordenamiento aplicable.

Un enfoque de derechos requiere evaluar de manera previa los sistemas empleados para monitorear espacios públicos, identificar usuarios o inferir opiniones políticas. Deben definirse la base legal, el objetivo,

el alcance temporal y personal, los mecanismos de autorización, los registros y los controles independientes. Las autoridades deben valorar si existen alternativas menos intrusivas y si el beneficio de la medida justifica la interferencia. La vigilancia indiscriminada o carente de salvaguardias puede comprometer más de un derecho a la vez.

La transparencia de los sistemas de recomendación y moderación debe ser suficiente para que los usuarios comprendan las reglas que afectan su participación. Esto no implica que cada decisión automatizada requiera una explicación técnica exhaustiva. Puede incluir información sobre la razón general de una medida, los criterios principales aplicados y el canal para solicitar revisión. La combinación de información, pluralidad de mecanismos de distribución y procedimientos de apelación contribuye a que las personas no queden sujetas a decisiones opacas que determinan su acceso al espacio público.

La protección de la libertad de expresión requiere, en suma, atender tanto a la supresión directa como a las formas menos visibles de reducción de alcance, priorización y vigilancia. Las garantías deben equilibrar los derechos de quienes publican, quienes reciben información y quienes pueden ser dañados por contenidos ilícitos. Ese equilibrio no se obtiene confiando exclusivamente en la plataforma ni prohibiendo de forma general la automatización. Requiere reglas claras, supervisión independiente y remedios accesibles.

3.4. Derechos laborales, sociales y de grupos vulnerables ante la IA

3.4.1. Trabajo, protección social y condiciones de vida

La incorporación de IA al trabajo puede automatizar tareas, complementar capacidades profesionales, modificar procesos y crear nuevas ocupaciones. Sus efectos dependen del sector, la organización, las tareas y las políticas adoptadas. La Organización Internacional del Trabajo (OIT), en su análisis global sobre IA generativa y empleo,

estimó que la tecnología tendría más probabilidades de transformar o complementar muchas ocupaciones que de sustituirlas por completo, aunque señaló la necesidad de políticas que protejan la calidad del empleo y faciliten transiciones justas. Esta conclusión no significa que ningún puesto vaya a desaparecer ni que las consecuencias sean iguales para todos los países o trabajadores.

Los derechos laborales pueden verse afectados por sistemas de selección, asignación de tareas, evaluación de desempeño, programación de horarios y vigilancia. Un modelo que filtra candidaturas puede excluir a personas por datos históricos o variables indirectas; una herramienta de productividad puede intensificar el ritmo de trabajo; y un sistema de gestión puede asignar turnos sin considerar responsabilidades de cuidado o necesidades razonables. La transparencia y la posibilidad de revisión son importantes, pero también lo son la consulta a trabajadores y sus representantes, la prevención de riesgos, la no discriminación y la protección de datos laborales conforme a las normas aplicables.

La gestión algorítmica puede modificar el equilibrio de poder entre empleador y trabajador. Si las decisiones sobre asignación de tareas, remuneración o sanciones se toman con base en puntuaciones opacas, el trabajador puede no saber cómo mejorar su situación o impugnar una evaluación. Las organizaciones deben determinar quién responde por la decisión, qué datos se usan y cómo se corrigen errores. La tecnología no debe impedir el ejercicio de derechos colectivos ni reemplazar los procedimientos de diálogo social y negociación reconocidos por el derecho laboral.

La IA también puede incidir en derechos sociales como educación, salud, vivienda y seguridad social. En salud, los sistemas pueden apoyar diagnóstico o priorización, pero un error puede afectar acceso a atención o continuidad de tratamiento. En educación, una herramienta puede personalizar contenidos, aunque también etiquetar a estudiantes o limitar expectativas con base en predicciones. En prestaciones

sociales, la automatización puede agilizar trámites, pero errores de identidad o de clasificación pueden suspender recursos esenciales. En todos estos sectores, la eficiencia debe compatibilizarse con revisión humana, accesibilidad, motivación y recursos efectivos.

La brecha digital puede convertir una innovación destinada a ampliar servicios en una nueva barrera. Cuando los trámites se trasladan a canales automatizados, las personas sin conexión estable, dispositivos, alfabetización digital o asistencia pueden quedar excluidas. La obligación de accesibilidad alcanza también a la información y a los sistemas de comunicación, y puede requerir formatos adaptados, tecnologías de asistencia y canales alternativos. La Convención sobre los Derechos de las Personas con Discapacidad reconoce la importancia de la accesibilidad a la información y las comunicaciones, incluidas las tecnologías y sistemas de información.

3.4.2. Protección de grupos vulnerables y salvaguardias diferenciadas

Las personas no se encuentran en la misma posición frente a las tecnologías. La edad, discapacidad, género, origen étnico, situación económica, idioma, condición migratoria y ubicación geográfica pueden influir en el acceso, la exposición al riesgo y la capacidad de reclamar. La vulnerabilidad no debe tratarse como una propiedad fija de ciertos grupos, sino como una condición que surge de relaciones de poder, barreras institucionales y circunstancias concretas. Las salvaguardias deben responder a esas condiciones y evitar que las etiquetas de vulnerabilidad reduzcan la agencia de las personas.

Las personas con discapacidad pueden beneficiarse de sistemas de reconocimiento de voz, subtitulado, lectura asistida o tecnologías de apoyo. También pueden verse perjudicadas si una herramienta interpreta de manera incorrecta patrones de comunicación, no es compatible con dispositivos de asistencia o impone un canal inaccesible. El diseño universal y la participación de personas con discapacidad desde las

etapas iniciales ayudan a evitar barreras. La accesibilidad no es un complemento opcional: es una condición para que puedan ejercer derechos en igualdad de condiciones.

Los niños, niñas y adolescentes requieren protección reforzada por su etapa de desarrollo y por la asimetría que existe entre ellos y las organizaciones que diseñan servicios digitales. Los sistemas de recomendación pueden influir en el acceso a contenidos y en la experiencia educativa; las herramientas de vigilancia pueden afectar privacidad y libertad; y los perfiles pueden permanecer vinculados a una persona durante largos periodos. Las evaluaciones deben considerar el interés superior del niño, su derecho a ser escuchado conforme a su edad y madurez, la protección de datos y la prevención de usos manipulativos o incompatibles con su desarrollo.

Las mujeres y personas de diversidades sexuales y de género pueden enfrentar riesgos particulares de discriminación, vigilancia o violencia facilitada por tecnología. Los modelos que aprenden de representaciones estereotipadas pueden reproducir exclusiones en selección de empleo o acceso a servicios. Los sistemas generativos pueden utilizarse para producir imágenes íntimas no consentidas, suplantaciones o campañas de acoso. La respuesta debe combinar prevención, mecanismos eficaces de denuncia y reparación, protección de víctimas y garantías de libertad de expresión, sin trasladar a las personas afectadas la carga de resolver por sí solas el daño.

Los pueblos indígenas y las comunidades que hablan lenguas con menor presencia digital pueden quedar subrepresentados en datos de entrenamiento y servicios automatizados. Una herramienta puede desconocer referencias culturales, clasificar incorrectamente información o excluir a una persona por expresarse en una lengua que el sistema procesa con menor calidad. La participación comunitaria, el respeto a derechos colectivos y el desarrollo de recursos lingüísticos adecuados pueden mejorar pertinencia y acceso. Las decisiones sobre

datos y sistemas que afectan a comunidades deben considerar quién controla la información y cómo se distribuyen los beneficios.

Las personas en situación de pobreza, desplazamiento o migración pueden depender en mayor medida de sistemas automatizados para acceder a ayudas, servicios o procedimientos administrativos. Una clasificación errónea puede ser especialmente grave cuando la persona no tiene asesoría, conectividad o documentación suficiente para impugnarla. Los Estados y las organizaciones deben ofrecer vías de atención humana, información clara y revisión sin barreras desproporcionadas. Una decisión automatizada no debe convertir la falta de recursos para reclamar en una aceptación tácita de su resultado.

La evaluación de impacto sobre derechos fundamentales debe prestar atención a estas diferencias. No basta con probar el modelo sobre un conjunto de datos promedio o evaluar su precisión técnica. Se deben identificar los grupos afectados, analizar posibles impactos diferenciados, consultar cuando corresponda, diseñar medidas de mitigación y establecer cómo se dará respuesta a incidentes. El enfoque debe ser continuo: cambios de modelo, datos, usuarios o contexto pueden modificar los riesgos iniciales.

Tabla 3.1

Derechos fundamentales, posibles afectaciones y salvaguardias

Derecho o ámbito protegido	Posible afectación relacionada con la IA	Salvaguardia jurídica o institucional
Dignidad humana	Reducir a una persona a una puntuación, perfil o categoría automatizada	Evaluar el propósito y los efectos del sistema; asegurar que la persona conserve reconocimiento y trato individual

Derecho o ámbito protegido	Posible afectación relacionada con la IA	Salvaguardia jurídica o institucional
Autonomía y libertad	Influir en decisiones personales mediante recomendaciones opacas o manipulativas	Informar sobre el uso del sistema y ofrecer opciones reales para aceptar, rechazar o cuestionar sus resultados
Igualdad y no discriminación	Producir errores o resultados desfavorables para determinados grupos	Evaluar impactos diferenciados, revisar datos y criterios, y corregir o retirar el sistema cuando corresponda
Libertad de expresión e información	Limitar o reducir la visibilidad de contenidos mediante moderación o clasificación automatizada	Comunicar los criterios relevantes y facilitar la impugnación de decisiones de moderación
Privacidad	Inferir aspectos personales, vigilar o clasificar a las personas sin garantías suficientes	Limitar la recopilación y el uso de datos, establecer controles de acceso y respetar los derechos de sus titulares
Trabajo y protección social	Automatizar selección, evaluación, asignación de tareas o acceso a prestaciones con efectos adversos	Informar a las personas afectadas, revisar decisiones relevantes y establecer supervisión y canales de reclamación
Derechos de personas y grupos en situación de vulnerabilidad	Excluir o perjudicar a personas con discapacidad, barreras lingüísticas o menor acceso digital	Diseñar sistemas accesibles, consultar a grupos afectados y mantener alternativas adecuadas de atención

Nota. *Las afectaciones y salvaguardias dependen del contexto, la finalidad y el derecho aplicable. Elaboración propia con base en los instrumentos de derechos humanos examinados en el capítulo, incluidos el Pacto Internacional de Derechos Civiles y Políticos, la Convención sobre los Derechos de las Personas con Discapacidad y la Convención Americana sobre Derechos Humanos.*

Capítulo 4.

Protección de datos personales, privacidad y decisiones automatizadas



INTRODUCCIÓN

Los sistemas de inteligencia artificial dependen con frecuencia de datos personales para clasificar, predecir, recomendar o generar resultados. Cuando estos datos permiten identificar a una persona, describen aspectos de su vida o se emplean para tomar decisiones que la afectan, entran en juego el derecho a la protección de datos y el derecho a la privacidad. Ambos se relacionan, pero no son idénticos: la privacidad protege ámbitos de la vida frente a injerencias indebidas, mientras que la protección de datos establece reglas para el tratamiento de información personal y reconoce facultades de control a sus titulares.

Este capítulo aborda cuatro cuestiones: la relación entre datos personales, privacidad y autodeterminación informativa; la elaboración de perfiles y las decisiones individuales automatizadas; el uso de biometría y la vigilancia algorítmica; y los derechos de los titulares y mecanismos de tutela. El análisis toma en cuenta el marco ecuatoriano, los estándares interamericanos y el Reglamento General de Protección de Datos de la Unión Europea, distinguiendo su ámbito territorial y fuerza jurídica. La comparación permite reconocer tendencias comunes sin asumir que una norma regional se aplica automáticamente en todos los países.

4.1. Datos personales, privacidad y autodeterminación informativa

4.1.1. Datos personales y alcance de la protección

Un dato personal es información que se refiere a una persona natural identificada o identificable. Puede identificarla de manera directa —por ejemplo, mediante su nombre o número de documento— o indirecta, cuando se combina con otros elementos que permiten distinguirla. La definición comprende más que los datos entregados voluntariamente por una persona: también puede incluir información observada, inferida o generada durante el uso de un servicio, siempre que se relacione con alguien identificable conforme a la legislación aplicable.

En un sistema de inteligencia artificial, los datos pueden aparecer en distintas fases. Durante la recopilación, pueden obtenerse de formularios, sensores, registros institucionales, plataformas o bases de terceros. En el entrenamiento, pueden utilizarse para ajustar el modelo o validar su rendimiento. En el despliegue, el sistema puede recibir información de una persona y producir una puntuación, predicción o recomendación que vuelva a vincularse con ella. El resultado inferido también puede ser dato personal cuando describe o permite tomar decisiones sobre una persona identificada o identificable. Por ello, la revisión de privacidad no debe limitarse a los datos iniciales.

La legislación ecuatoriana reconoce la protección de datos personales como un derecho constitucional. El artículo 66, numeral 19, de la Constitución de la República del Ecuador incluye el acceso y la decisión sobre información y datos de ese carácter, así como su protección. La Ley Orgánica de Protección de Datos Personales (LOPDP), publicada en el Quinto Suplemento del Registro Oficial No. 459 el 26 de mayo de 2021, desarrolla ese derecho y regula el tratamiento automatizado y no automatizado de datos personales. Su aplicación debe analizarse junto con el Reglamento General y las normas emitidas por la autoridad de control.

No toda información vinculada a una persona recibe el mismo nivel de protección, aunque todo dato personal debe tratarse conforme a los principios y bases jurídicas correspondientes. Algunas categorías —entre ellas las relativas a salud, origen étnico, orientación sexual, convicciones y biometría— pueden requerir protección reforzada, especialmente cuando su uso indebido puede generar discriminación o afectar derechos fundamentales. La LOPDP ecuatoriana incluye datos biométricos dentro de las categorías de datos sensibles. Esto no significa que toda imagen facial tenga automáticamente el mismo tratamiento que un patrón biométrico: importa cómo se procesa y si el dato permite o confirma la identificación única de una persona.

El principio de finalidad exige que la información se recoja y utilice para propósitos determinados, explícitos y legítimos, conforme a la base jurídica aplicable. Una entidad que obtiene datos para prestar un servicio no puede presumir que puede reutilizarlos para cualquier análisis predictivo o perfil comercial. La compatibilidad de un uso ulterior debe evaluarse a la luz de la ley, la finalidad original, las expectativas razonables de las personas y los riesgos que genera el nuevo tratamiento. La posibilidad técnica de combinar bases de datos no equivale a una autorización jurídica para hacerlo.

Otros principios relevantes son la minimización, la exactitud, la limitación de conservación, la seguridad, la transparencia y la responsabilidad proactiva. La minimización obliga a valorar si cada dato es necesario para la finalidad; la exactitud exige mecanismos de actualización y corrección; y la limitación de conservación requiere definir plazos adecuados, no guardar información indefinidamente por si pudiera ser útil. En modelos de aprendizaje automático, estos principios pueden ser difíciles de aplicar cuando los datos se replican en conjuntos de entrenamiento, registros intermedios o copias de respaldo. Esa complejidad exige controles técnicos y organizativos que permitan localizar, proteger y, cuando corresponda, rectificar o suprimir información.

La información puede ser seudonimizada, es decir, separada de identificadores directos mediante claves o códigos. Esto reduce ciertos riesgos, pero no siempre la convierte en anónima: si es razonablemente posible volver a identificar a la persona usando información adicional, continúa siendo dato personal conforme al régimen aplicable. La anonimización efectiva, en cambio, requiere que la reidentificación no sea razonablemente posible considerando medios, contexto y tecnología disponibles. La afirmación de una organización de que "los datos no tienen nombres" no basta para concluir que están fuera del ámbito de protección.

La integración de datos aumenta la capacidad de inferencia. Registros aparentemente ordinarios pueden revelar hábitos, ubicación, estado de salud o preferencias al combinarse. El análisis jurídico debe valorar no solo cada campo aislado, sino también la capacidad del conjunto para reconstruir aspectos íntimos. Esta dimensión es especialmente relevante en IA, donde un modelo puede inferir características que la persona nunca proporcionó de forma directa.

4.1.2. Privacidad y autodeterminación informativa

El derecho a la privacidad protege a las personas frente a injerencias arbitrarias o ilegales en su vida personal y familiar, domicilio, comunicaciones y otros ámbitos protegidos. El derecho a la protección de datos aporta reglas más específicas sobre la obtención, uso, conservación y circulación de información personal. Una práctica puede plantear problemas de privacidad aunque no viole una regla técnica de seguridad; también puede incumplir principios de protección de datos sin que exista una divulgación pública. Ambos derechos se complementan en la evaluación de la IA.

La autodeterminación informativa alude a la capacidad de la persona para conocer y ejercer control sobre el tratamiento de información que la concierne. No significa que cada titular pueda determinar individualmente todos los usos de cada dato en cualquier circunstancia.

Los ordenamientos prevén bases jurídicas distintas del consentimiento y pueden establecer excepciones. La idea central es que el tratamiento debe tener fundamento, finalidad y límites, y que las personas deben contar con derechos y garantías frente a usos que las afecten.

El consentimiento es una base importante en ciertos contextos, pero no debe convertirse en una respuesta automática para legitimar cualquier tratamiento. Debe reunir las condiciones establecidas en la ley —como libertad, información y especificidad— y puede ser insuficiente cuando existe un desequilibrio de poder o cuando la persona no tiene una alternativa real. En servicios esenciales, relaciones laborales o trámites públicos, aceptar condiciones extensas puede no reflejar una elección genuina. En esos escenarios, la entidad debe identificar la base jurídica adecuada y justificar la necesidad y proporcionalidad del tratamiento.

La transparencia debe informar, entre otros aspectos, quién trata los datos, para qué, con qué base jurídica, cuánto tiempo se conservarán, si se comunicarán a terceros y qué derechos pueden ejercerse. Cuando hay decisiones automatizadas o elaboración de perfiles, la información debe permitir comprender que tales procesos existen y, en los casos previstos por la ley, cuáles son sus consecuencias relevantes. La explicación no necesita exponer todo el código o revelar secretos comerciales; debe permitir que el titular entienda de forma práctica qué tratamiento se realiza y cómo puede ejercer sus derechos.

En América, los Principios actualizados sobre privacidad y protección de datos personales del Comité Jurídico Interamericano de la Organización de los Estados Americanos (OEA) ofrecen una referencia regional. Abordan finalidades lícitas, transparencia y consentimiento, pertinencia y necesidad, derechos de acceso, rectificación, supresión y oposición, entre otros aspectos. La OEA los presenta como un instrumento de *soft law* para orientar el fortalecimiento de marcos nacionales, no como una ley regional directamente aplicable en todos los Estados.

La protección desde el diseño requiere que las medidas de privacidad se incorporen antes de desplegar un sistema, en lugar de añadirse únicamente después de una queja o una filtración. Puede implicar recolectar menos datos, separar identificadores, limitar el acceso, fijar plazos de conservación y diseñar interfaces que faciliten el ejercicio de derechos. La Superintendencia de Protección de Datos Personales del Ecuador ha publicado una guía sobre protección de datos desde el diseño y por defecto, que sirve como referencia administrativa en el contexto nacional.

Cuando el tratamiento puede generar un riesgo alto para los derechos y libertades de las personas, la evaluación de impacto permite anticipar riesgos, justificar la necesidad del sistema y documentar medidas de mitigación. El Reglamento General de la LOPDP prevé evaluaciones de impacto obligatorias en los casos determinados por la ley y desarrolla criterios vinculados a tratamientos automatizados que evalúan sistemáticamente aspectos personales, al tratamiento a gran escala y a otros supuestos de riesgo. Esta obligación debe aplicarse según los hechos y la normativa ecuatoriana vigente, no suponerse por el mero hecho de que se emplee una herramienta de IA.

Tabla 4.1

Etapas del tratamiento de datos y controles de protección

Etapa	Riesgo relacionado con la IA	Control de protección
Recolección de datos	Obtener más información de la necesaria o sin fundamento jurídico adecuado	Definir la finalidad, identificar la base jurídica aplicable y limitar los datos a los necesarios
Preparación y selección	Usar datos incompletos, inexactos o poco representativos	Verificar su calidad, pertinencia, procedencia y actualización

Etapa	Riesgo relacionado con la IA	Control de protección
Análisis y elaboración de perfiles	Inferir características personales sensibles o producir perfiles incorrectos	Evaluar la necesidad y proporcionalidad del perfilamiento; revisar la validez de las inferencias
Toma de decisiones automatizadas	Adoptar decisiones que afecten significativamente a una persona sin garantías suficientes	Informar sobre el uso del sistema y habilitar revisión humana y vías de impugnación cuando correspondan
Conservación	Mantener datos por más tiempo del necesario o sin controles de seguridad	Establecer plazos de conservación, medidas de seguridad y procedimientos de eliminación
Comunicación o transferencia	Compartir datos con proveedores o terceros sin controles adecuados	Definir responsabilidades, limitar accesos y verificar las condiciones aplicables a la transferencia
Ejercicio de derechos y reclamación	Dificultar que una persona conozca, corrija o cuestione el tratamiento	Ofrecer canales accesibles para solicitudes, revisión, reclamación y respuesta

Nota. *Los controles deben ajustarse a la finalidad, el tipo de datos, el contexto y los riesgos del tratamiento. Las evaluaciones de impacto se realizan en los casos previstos por la normativa aplicable. Elaboración propia con base en la Ley Orgánica de Protección de Datos Personales de Ecuador y el Reglamento General de Protección de Datos de la Unión Europea.*

4.2. Elaboración de perfiles y decisiones individuales automatizadas

4.2.1. Perfilamiento y usos de las inferencias

La elaboración de perfiles consiste, en términos generales, en tratar datos personales para evaluar o predecir aspectos relativos a una persona, como su situación económica, preferencias, intereses, comportamiento o ubicación. La técnica puede ser útil para adaptar servicios, detectar

necesidades o identificar riesgos. También puede generar categorías que influyen en el trato que recibe alguien, a veces sin que conozca la existencia del perfil o pueda verificar la información que lo compone.

Un perfil no es una descripción neutral e inmutable de la persona. Es el resultado de decisiones sobre datos, variables, objetivos, categorías y umbrales. Puede estar basado en información correcta y, aun así, producir inferencias discutibles. Un sistema que predice una conducta con cierta probabilidad no demuestra que esa conducta haya ocurrido ni que vaya a ocurrir. La predicción describe una estimación bajo condiciones determinadas, no una certeza ni una característica esencial del titular.

Las consecuencias pueden aparecer en ámbitos diversos. Un perfil puede influir en la oferta de productos, el precio, el acceso a una oportunidad laboral, la priorización de una solicitud, la selección de contenido o la clasificación de un caso para una revisión adicional. El riesgo aumenta cuando una inferencia se utiliza para adoptar medidas con efectos jurídicos o consecuencias significativas, como denegar un servicio, alterar condiciones contractuales o afectar el acceso a una prestación. En esos casos, la entidad debe evaluar tanto la base legal del tratamiento como las garantías aplicables al proceso decisorio.

El perfilamiento puede utilizar datos proporcionados, observados o inferidos. La información observada puede incluir señales de navegación o ubicación; la inferida puede atribuir preferencias, nivel de riesgo o probabilidad de impago. Una persona puede no saber qué factores influyeron en la inferencia ni cómo corregirla. La exactitud del dato original tampoco garantiza la exactitud del perfil: la relación estadística puede no ser adecuada para la finalidad concreta o puede variar entre poblaciones.

Por ello, el responsable del tratamiento debe documentar la finalidad del perfil, las fuentes de datos, las categorías de información utilizadas,

los destinatarios, los plazos de conservación y las consecuencias previstas. También debe evaluar si el perfilamiento es necesario, si puede lograrse el objetivo con menos datos y si el titular fue informado de forma comprensible. Cuando la decisión depende de información incorrecta, desactualizada o fuera de contexto, la posibilidad de rectificación adquiere especial importancia.

4.2.2. Decisiones automatizadas: alcance y salvaguardias

Una decisión automatizada es aquella en la que un sistema participa en la determinación de un resultado respecto de una persona. El grado de automatización puede variar: la herramienta puede limitarse a ordenar información, formular una recomendación o producir una salida que se convierte directamente en la decisión final. La existencia de una firma humana al final del proceso no demuestra necesariamente que la decisión haya sido revisada de manera sustantiva. Debe evaluarse cuánto influye la salida y qué capacidad real tiene la persona responsable para apartarse de ella.

El Reglamento General de Protección de Datos de la Unión Europea (RGPD) reconoce en su artículo 22 el derecho a no quedar sujeto a una decisión basada únicamente en tratamiento automatizado, incluida la elaboración de perfiles, cuando produce efectos jurídicos o afecta de modo similar significativamente a la persona. El artículo contempla excepciones específicas —por ejemplo, en ciertos supuestos de necesidad contractual, autorización legal o consentimiento explícito— y establece salvaguardias aplicables, como la posibilidad de obtener intervención humana, expresar el propio punto de vista e impugnar la decisión, según el caso. Esta disposición tiene alcance europeo conforme al RGPD y no debe presentarse como regla automáticamente vigente en Ecuador.

La LOPDP ecuatoriana formula una protección propia. Su artículo 20 reconoce el derecho a no ser sometido a una decisión basada única o parcialmente en valoraciones producto de procesos automatizados,

incluida la elaboración de perfiles, cuando produzca efectos jurídicos o atente contra derechos y libertades fundamentales. Entre las facultades previstas se encuentran solicitar una explicación motivada, presentar observaciones, requerir información sobre los criterios de valoración, conocer los tipos de datos usados y su fuente, e impugnar la decisión. La mención de decisiones basadas "única o parcialmente" hace necesario considerar situaciones en las que la intervención humana no elimina el peso determinante del perfil automatizado.

La aplicación de estas garantías requiere distinguir entre un apoyo técnico y una decisión que afecta derechos. Un sistema que ordena expedientes para que un funcionario los revise puede tener una influencia menor que otro que determina automáticamente el acceso a una prestación. Pero incluso una priorización puede producir efectos significativos si los casos relegados reciben atención tarde o nunca. La evaluación debe atender al funcionamiento real del procedimiento, los efectos previsibles y la posibilidad de revisión, no solo a la etiqueta administrativa que se da al sistema.

Una explicación significativa debe ser comprensible y útil para el ejercicio de derechos. En el caso C-203/22, *CK contra Magistrat der Stadt Wien*, el Tribunal de Justicia de la Unión Europea abordó el acceso a información sobre la lógica de un perfil utilizado en la evaluación de solvencia. El Tribunal señaló que una fórmula matemática compleja o una descripción exhaustiva de pasos no bastan por sí solas si no permiten comprender qué datos se utilizaron y cómo influyeron en la decisión. La información debe permitir una comprensión efectiva, sin que la complejidad técnica elimine el deber de explicar. Esta interpretación corresponde al RGPD en el sistema jurídico europeo, pero resulta relevante como referencia comparada sobre transparencia práctica.

El secreto comercial no elimina automáticamente los derechos del titular. Puede justificar que ciertos elementos no se divulguen al

público, pero las autoridades o tribunales pueden necesitar acceder a información para supervisar la legalidad y resolver una impugnación. Deben buscarse mecanismos que protejan intereses legítimos sin impedir el control, como acceso confidencial, peritajes independientes o explicaciones que revelen los factores decisivos sin publicar el código fuente.

4.2.3. Automatización y evaluación de riesgos

Antes de usar perfiles en decisiones relevantes, la organización debe identificar quién define el objetivo, qué grupos podrían verse afectados, qué datos se emplean y qué margen de error se considera aceptable. También debe comprobar si el sistema funciona en el contexto concreto, porque su rendimiento puede variar cuando cambian los datos, el idioma, la población o las condiciones institucionales. Una validación general del proveedor no sustituye la evaluación contextual de quien despliega la herramienta.

La revisión humana debe ser significativa. La persona que revisa necesita acceso a los factores relevantes, formación suficiente, tiempo para examinar el expediente y autoridad para modificar o rechazar la recomendación. Si se espera que confirme todas las salidas por razones de productividad, la supervisión se convierte en una formalidad. La entidad debe documentar los criterios de revisión y conservar información que permita reconstruir por qué se adoptó la decisión.

La evaluación debe incluir vías de reclamación y corrección. La persona necesita saber a quién dirigirse, cómo presentar información adicional y en qué plazo se revisará su caso. Si el sistema utiliza datos de terceros, la organización debe determinar cómo se verifica la fuente y cómo se corrige un dato inexacto. La reparación puede requerir modificar el resultado individual, detener temporalmente el uso del modelo o revisar decisiones similares que afectaron a otras personas.

Una decisión automatizada no se vuelve legítima únicamente porque el modelo tenga alta precisión. La entidad debe demostrar que el uso es pertinente, legal y proporcionado a la finalidad; que no hay una alternativa menos invasiva igualmente eficaz; y que los derechos del titular se protegen mediante controles adecuados. Esta valoración adquiere mayor importancia cuando el tratamiento incluye categorías sensibles, afecta a niños, niñas y adolescentes o puede generar consecuencias graves.

4.3. Biometría, reconocimiento facial y vigilancia algorítmica

4.3.1. Datos biométricos e identificación

Los datos biométricos se derivan de características físicas, fisiológicas o conductuales de una persona y pueden permitir o confirmar su identificación única. Entre los ejemplos se encuentran huellas dactilares, patrones faciales, iris, voz o formas de movimiento, según la técnica y la finalidad. No toda fotografía es automáticamente un dato biométrico utilizado para identificación única. Sin embargo, cuando se somete a un tratamiento técnico destinado a identificar o autenticar a una persona, puede adquirir esa naturaleza bajo los marcos aplicables.

La biometría presenta riesgos particulares porque ciertas características son persistentes y difíciles de sustituir. Una contraseña comprometida puede cambiarse; un rostro o una huella no se reemplazan de la misma manera. Una filtración puede permitir usos posteriores no previstos, suplantación o identificación en distintos contextos. Por ello, las organizaciones deben limitar la recolección y conservación, proteger las plantillas biométricas, controlar accesos y evitar que una finalidad —por ejemplo, verificar identidad— se amplíe a otra incompatible, como seguimiento continuo.

El RGPD considera categoría especial los datos biométricos tratados con el fin de identificar de manera unívoca a una persona, sujetos a las condiciones previstas en su artículo 9. En Ecuador, la LOPDP

incorpora los datos biométricos entre los datos sensibles y establece condiciones específicas para el tratamiento de categorías especiales. La Superintendencia de Protección de Datos Personales ha señalado, en consultas institucionales, que las fotografías y huellas pueden constituir datos personales especiales de naturaleza biométrica, según el tratamiento y la definición legal aplicable. La existencia de una imagen facial, por sí sola, debe distinguirse del proceso técnico de reconocimiento o identificación.

La biometría puede emplearse para desbloquear dispositivos, controlar acceso a instalaciones, verificar identidad en trámites o identificar personas en espacios públicos. La finalidad y el contexto cambian la intensidad de la injerencia. Una verificación puntual, voluntaria y con alternativa accesible no plantea necesariamente los mismos riesgos que la identificación remota de personas que transitan en un espacio público. En este último caso, pueden verse afectadas personas que no han consentido participar, incluidas quienes no son objeto de investigación o sospecha.

El uso de datos biométricos debe justificarse conforme a la base jurídica y a los principios de finalidad, necesidad y proporcionalidad. El consentimiento no siempre es una base suficiente, particularmente si la persona no tiene una alternativa real o si el tratamiento se realiza por una entidad pública. La institución debe determinar si existe una habilitación legal adecuada, qué garantías exige y si la medida es necesaria para cumplir un objetivo legítimo. La autoridad ecuatoriana ha advertido que una disposición que permite ciertos tratamientos no debe leerse automáticamente como autorización general para usar biometría en todo contexto laboral o público.

4.3.2. Reconocimiento facial y errores de identificación

Los sistemas de reconocimiento facial pueden verificar si dos imágenes corresponden a la misma persona o buscar una coincidencia dentro de una base de datos. La verificación suele comparar una muestra con

una identidad declarada; la identificación busca determinar quién es una persona entre varias identidades posibles. Aunque ambas técnicas empleen imágenes faciales, sus finalidades y riesgos difieren. La identificación masiva puede capturar a muchas personas sin que estén informadas o involucradas en una investigación, lo que amplía el alcance de la vigilancia.

Los errores pueden producir falsos positivos —atribuir a una persona la identidad de otra— y falsos negativos —no reconocer correctamente a alguien—. Sus efectos dependen del uso. En un sistema de acceso, un error puede impedir el ingreso; en seguridad pública, puede conducir a una detención, investigación o registro indebido. El riesgo aumenta si la decisión se adopta sin verificación humana, si la base de referencia contiene datos de baja calidad o si el sistema funciona peor para algunos grupos.

La tasa de error no debe evaluarse solo en promedio. Es necesario examinarla por grupos, condiciones de captura, iluminación, edad, tono de piel y otros factores pertinentes, además de comprobar que las pruebas representen el contexto real. La evaluación técnica debe complementarse con controles institucionales: quién autoriza la búsqueda, qué umbral de coincidencia se utiliza, qué evidencia adicional se requiere y cómo se corrige una identificación errónea. Una coincidencia algorítmica es una pista o resultado técnico, no una prueba autónoma de identidad o responsabilidad.

Las imágenes faciales pueden recogerse de cámaras, redes sociales, documentos de identidad u otras bases. Que una imagen esté disponible en internet no significa que su utilización para reconocimiento biométrico sea automáticamente lícita. Deben analizarse la finalidad original, la base jurídica, la expectativa razonable, las reglas sobre datos sensibles y las limitaciones aplicables. La extracción masiva de imágenes para construir bases de reconocimiento puede afectar a personas que nunca tuvieron oportunidad de conocer el uso o ejercer

sus derechos.

4.3.3. Vigilancia algorítmica y salvaguardias

La vigilancia algorítmica combina datos, herramientas de análisis y sistemas de clasificación para observar conductas, identificar patrones o predecir riesgos. Puede integrar reconocimiento facial, geolocalización, redes de cámaras, análisis de redes sociales, sensores y bases administrativas. La combinación de fuentes puede revelar movimientos, asociaciones, rutinas y relaciones. Por ello, los riesgos no se limitan a la privacidad individual: pueden afectar libertad de expresión, reunión, circulación, igualdad y participación política.

La evaluación jurídica debe considerar la base legal, la finalidad, el alcance personal y territorial, la duración, las personas con acceso, los criterios de retención y los mecanismos de supervisión. También debe analizar si existen alternativas menos intrusivas y si la vigilancia es necesaria y proporcional. Una autorización general para mantener el orden no necesariamente habilita cualquier técnica de seguimiento masivo. Las restricciones deben establecerse con suficiente claridad y someterse a control independiente.

La vigilancia puede producir un efecto inhibidor. Una persona que sabe o sospecha que sus comunicaciones, desplazamientos o contactos son analizados puede evitar asistir a una protesta, hablar con periodistas o participar en actividades públicas. Este efecto puede afectar especialmente a defensores de derechos humanos, minorías, periodistas y comunidades ya sujetas a vigilancia desproporcionada. La ausencia de una sanción inmediata no elimina el impacto sobre el ejercicio de derechos.

Las medidas de salvaguardia incluyen evaluaciones de impacto, controles de acceso, registros de consultas, límites de conservación, auditorías independientes, supervisión judicial o administrativa

y procedimientos de notificación cuando sea compatible con la investigación legítima. En ciertos usos, la entidad puede necesitar mecanismos de revisión antes de activar la identificación o de actuar sobre una coincidencia. Deben existir reglas para suspender el sistema ante fallos, sesgos o uso fuera de la finalidad autorizada.

La autoridad de protección de datos puede desempeñar un papel relevante al supervisar tratamientos, emitir guías, recibir reclamaciones y exigir medidas correctivas según su competencia. En Ecuador, la LOPDP y el Reglamento General establecen el marco nacional, mientras que las disposiciones concretas de la Superintendencia deben identificarse por su naturaleza y vigencia. En el plano regional, los Principios de la OEA ofrecen orientación para fortalecer la protección de datos y privacidad, pero no sustituyen las normas vinculantes de cada país.

4.4. Derechos de los titulares y mecanismos de tutela en entornos de IA

4.4.1. Ejercicio de derechos y deberes del responsable

Los derechos del titular permiten conocer y controlar, dentro de los límites legales, el tratamiento de sus datos. En Ecuador, la LOPDP contempla derechos como información, acceso, rectificación y actualización, eliminación, oposición, portabilidad, suspensión o limitación del tratamiento en los casos previstos, y el derecho a no ser objeto de determinadas decisiones basadas en valoraciones automatizadas. El ejercicio práctico requiere que la persona sepa quién es el responsable y disponga de canales accesibles para presentar solicitudes.

El derecho de información permite conocer las finalidades, la base legal, los tipos de tratamiento, los plazos de conservación y la existencia de decisiones automatizadas cuando corresponda. El derecho de acceso facilita obtener confirmación sobre el tratamiento y acceder a los datos personales. La rectificación y actualización

permiten corregir información incorrecta o incompleta. La eliminación puede proceder bajo supuestos legales, aunque puede estar sujeta a excepciones, por ejemplo, cuando la conservación sea necesaria para cumplir una obligación legal o atender reclamaciones. La oposición y limitación permiten solicitar que el tratamiento cese o se restrinja en las circunstancias previstas por la ley.

El derecho a la portabilidad puede facilitar que los datos se transmitan a otro responsable en los casos y condiciones establecidos. En entornos de IA, su ejercicio puede ser más complejo si los datos se encuentran integrados en modelos, perfiles y sistemas de terceros. El responsable debe contar con procedimientos para responder de manera razonable y explicar si una solicitud no puede satisfacerse íntegramente. La dificultad técnica no elimina por sí sola una obligación legal; puede requerir medidas de diseño que hagan posible localizar, corregir o restringir datos.

El responsable del tratamiento debe identificar una base legal adecuada y respetar los principios durante todo el ciclo de vida. Si contrata a un proveedor que procesa datos por su cuenta, debe regular la relación y verificar que existan garantías suficientes. El proveedor, por su parte, debe cumplir las obligaciones que le correspondan como encargado o responsable, según sus decisiones sobre finalidades y medios. La clasificación contractual no debe usarse para ocultar quién determina realmente el tratamiento.

La protección de datos desde el diseño y por defecto implica incorporar salvaguardias en la arquitectura del sistema y en su configuración inicial. En la práctica, puede traducirse en recolectar solo los datos necesarios, establecer perfiles de acceso, limitar exportaciones, reducir la retención, impedir usos secundarios incompatibles y ofrecer opciones de privacidad comprensibles. La guía de la Superintendencia ecuatoriana sobre esta materia aporta orientación para organizaciones sujetas a la LOPDP.

La evaluación de impacto también puede ordenar la gestión de un sistema de IA. Debe describir el tratamiento, valorar su necesidad y proporcionalidad, identificar riesgos para los derechos y documentar medidas para reducirlos. No debería limitarse a completar un formulario previo al despliegue: si cambian el modelo, los datos, la finalidad o la población afectada, la evaluación puede necesitar revisión. En Ecuador, el Reglamento General dispone que las evaluaciones obligatorias se realicen antes de iniciar el tratamiento en los casos previstos.

4.4.2. Reclamación, supervisión y reparación

Cuando el titular considera que sus derechos han sido vulnerados, la primera vía suele ser la solicitud o reclamación ante el responsable del tratamiento, conforme al procedimiento aplicable. Si la respuesta no es satisfactoria o la ley lo permite, puede acudirse a la autoridad de protección de datos y, según el caso, a la vía judicial. La persona debe poder identificar plazos, requisitos y canales sin enfrentarse a obstáculos que hagan ilusorio el derecho. La información técnica no debe ser una condición imposible para presentar una queja.

La autoridad de control puede investigar el tratamiento, solicitar documentación, evaluar medidas de seguridad y disponer acciones correctivas dentro de sus competencias. El control es importante cuando la persona no tiene acceso al código, a los datos de entrenamiento o a los contratos con proveedores. La autoridad puede revisar información que no sea públicamente divulgable y determinar si el sistema cumple las obligaciones legales, respetando al mismo tiempo garantías procedimentales y la confidencialidad legítima.

En una controversia sobre decisión automatizada, la persona puede necesitar conocer los datos usados, la lógica esencial, los factores que influyeron y las consecuencias previstas. No siempre será necesario revelar el algoritmo íntegro. La información debe bastar para comprender el papel del tratamiento, verificar la exactitud de los datos y ejercer el derecho de impugnación. El Tribunal de Justicia de

la Unión Europea, en el asunto C-203/22, precisó que la complejidad matemática no libera al responsable de proporcionar información comprensible sobre los datos utilizados y la forma en que influyeron en la decisión, conforme al RGPD. Este criterio es relevante como referencia interpretativa europea, mientras que las reclamaciones en Ecuador deben resolverse con base en la LOPDP y la normativa nacional.

La tutela puede requerir medidas distintas según el daño. Una corrección de datos puede ser suficiente en un caso; en otro, será necesario reconsiderar la decisión, restituir una prestación, indemnizar daños o suspender el sistema. También puede haber afectaciones colectivas cuando el mismo modelo perjudica a un grupo de personas. La revisión de un caso individual no siempre resuelve el patrón sistémico, por lo que la autoridad o el tribunal puede necesitar examinar la práctica general y ordenar cambios organizativos o técnicos.

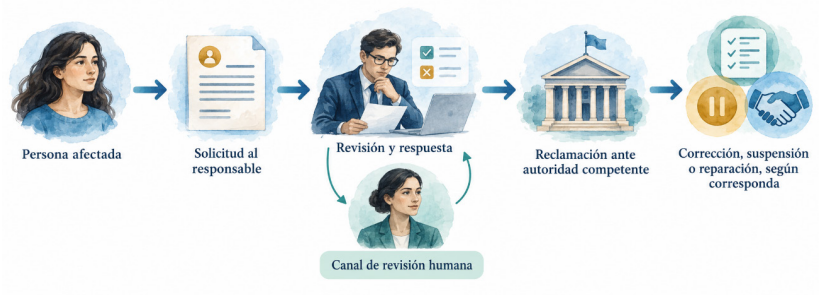
La prueba plantea dificultades cuando el funcionamiento está distribuido entre varios actores o cuando los registros no se conservaron. Resulta importante que existan documentación, registros de versiones, trazabilidad de decisiones y contratos claros sobre acceso a información y cooperación en auditorías. El responsable debe poder demostrar cumplimiento conforme a los deberes de responsabilidad proactiva. La falta de transparencia no debería hacer imposible que el titular conozca los hechos esenciales o que una autoridad investigue.

Los mecanismos de tutela deben proteger también a personas que enfrentan barreras lingüísticas, económicas, digitales o de discapacidad. Una plataforma exclusivamente en línea puede dificultar el reclamo de quienes no tienen acceso estable a internet; un formulario técnico puede ser incomprensible; y un plazo breve puede ser inadecuado si no se informó de la decisión. La accesibilidad del procedimiento forma parte de la protección efectiva de los derechos.

La comparación entre Ecuador y la Unión Europea muestra elementos comunes —transparencia, control del titular y garantías ante decisiones automatizadas— y diferencias en formulación, ámbito y mecanismos. La LOPDP ecuatoriana menciona decisiones basadas única o parcialmente en valoraciones automatizadas y reconoce facultades específicas de explicación, observación e impugnación. El RGPD se centra, en su artículo 22, en decisiones basadas únicamente en tratamiento automatizado que produzcan efectos jurídicos o similares significativos, con excepciones y salvaguardias. Ninguno de estos marcos debe aplicarse fuera de su jurisdicción sin analizar las reglas territoriales correspondientes.

Figura 4.1.

Ruta para ejercer derechos frente a una decisión automatizada



Nota: *Elaboración propia. No presentar los plazos ni resultados del proceso como iguales en todos los países.*



Capítulo 5.

Regulación internacional y comparada de la inteligencia artificial



INTRODUCCIÓN

La regulación de la inteligencia artificial se desarrolla mediante normas jurídicas, políticas públicas, estándares técnicos y acuerdos internacionales que no tienen la misma fuerza ni el mismo alcance. Algunos ordenamientos establecen obligaciones generales para los sistemas de IA; otros regulan usos concretos mediante normas sectoriales, reglas sobre datos personales o medidas de seguridad y contenidos. También existen recomendaciones y marcos voluntarios que orientan la conducta sin crear, por sí solos, obligaciones jurídicas directamente exigibles.

Este capítulo compara el Reglamento de Inteligencia Artificial de la Unión Europea, los enfoques de Estados Unidos y China y algunas experiencias de otras jurisdicciones. Después examina los instrumentos de organismos internacionales y el papel de la convergencia normativa, el denominado efecto Bruselas y la competencia regulatoria. El análisis es una fotografía jurídica al **20 de octubre de 2026**. Esta precisión importa porque los proyectos pueden convertirse en leyes, las normas pueden reformarse y los tratados solo obligan a los Estados conforme a sus reglas de firma, ratificación y entrada en vigor.

5.1. El Reglamento de IA de la Unión Europea y el enfoque basado en riesgos

5.1.1. Estructura y ámbito del Reglamento

El Reglamento (UE) 2024/1689, conocido como Reglamento de Inteligencia Artificial o *AI Act*, es el principal marco horizontal de la Unión Europea en esta materia. Su finalidad es establecer reglas armonizadas para el desarrollo, comercialización, puesta en servicio y uso de sistemas de IA en el mercado europeo, promoviendo al mismo tiempo la innovación y la protección de los derechos y valores reconocidos en la Unión. Es un reglamento de alcance regional: su aplicación debe determinarse conforme a sus disposiciones sobre ámbito, sujetos y efectos territoriales, y no debe confundirse con una ley mundial.

El Reglamento emplea una definición funcional de sistema de IA y organiza las obligaciones de acuerdo con el tipo de práctica y el nivel de riesgo. Esta estructura permite que una misma tecnología quede sujeta a obligaciones distintas según la finalidad y el contexto. Un modelo utilizado para una tarea interna de bajo impacto no recibe necesariamente el mismo tratamiento que un sistema empleado para decidir el acceso a empleo, educación, servicios esenciales o determinados procedimientos públicos.

El ámbito territorial puede alcanzar a proveedores, importadores, distribuidores y responsables del despliegue, incluso cuando se encuentren fuera de la Unión, si se cumplen las condiciones previstas en el Reglamento, por ejemplo, cuando el sistema se comercializa o se utiliza en la Unión o cuando sus resultados se usan allí. Esta dimensión extraterritorial ayuda a explicar por qué empresas de otros países pueden necesitar adaptar sus prácticas al marco europeo para operar en ese mercado. La obligación concreta, sin embargo, debe analizarse caso por caso y no inferirse simplemente por la presencia de usuarios europeos.

El Reglamento no reemplaza el resto del derecho de la Unión. Las obligaciones de protección de datos, consumo, igualdad, seguridad de productos, competencia y derecho laboral pueden continuar aplicándose de manera concurrente. Si un sistema procesa datos personales, por ejemplo, el cumplimiento del Reglamento de IA no elimina las exigencias del Reglamento General de Protección de Datos. La conformidad técnica tampoco resuelve automáticamente si una decisión es proporcional, si existe discriminación o si una persona tuvo acceso a una vía efectiva de impugnación.

La clasificación por riesgos se apoya en prohibiciones, obligaciones para determinados sistemas de alto riesgo y requisitos de transparencia para usos específicos. También incorpora reglas para modelos de IA de propósito general y mecanismos institucionales de supervisión y apoyo. En vez de imponer una única carga a todos los sistemas, el enfoque vincula la intensidad de las obligaciones con los riesgos que el legislador considera relevantes para la salud, la seguridad y los derechos fundamentales.

5.1.2. Categorías de riesgo y obligaciones principales

El Reglamento prohíbe ciertas prácticas que considera incompatibles con los valores y derechos de la Unión, con las excepciones y condiciones específicas establecidas en el propio texto. Entre ellas se encuentran determinadas formas de manipulación dañina, explotación de vulnerabilidades y puntuación social; también restringe ciertos usos de identificación biométrica y categorización. La calificación precisa de una práctica como prohibida depende de su definición normativa, su finalidad, sus efectos y las excepciones previstas. No basta con que una herramienta emplee una técnica similar para concluir que está automáticamente prohibida.

Los sistemas de alto riesgo ocupan un lugar central. El Reglamento identifica ámbitos y usos en los que los sistemas pueden afectar significativamente la vida de las personas, como empleo, educación,

acceso a servicios esenciales, ciertos usos de seguridad y justicia, infraestructuras críticas y migración. Para esos sistemas contempla requisitos relativos a la gestión de riesgos, gobernanza de datos, documentación técnica, registros, transparencia e información a los responsables del despliegue, supervisión humana, precisión, robustez y ciberseguridad. La lista y las condiciones deben leerse conforme a los anexos y artículos aplicables.

La lógica de alto riesgo no supone que el uso sea ilícito por naturaleza. Significa que puede realizarse sujeto a requisitos de conformidad y responsabilidades específicas. El proveedor tiene deberes propios de diseño, gestión de riesgos y documentación; el responsable del despliegue debe utilizar el sistema conforme a instrucciones y establecer supervisión adecuada; y otros actores pueden tener obligaciones según su papel y las circunstancias. La asignación precisa depende de la actividad, de la cadena de suministro y de las disposiciones del Reglamento.

La categoría de riesgo limitado incorpora obligaciones de transparencia para ciertos usos. Por ejemplo, las personas deben recibir información en situaciones reguladas en las que interactúan con un sistema de IA o se exponen a determinados contenidos sintéticos. Las obligaciones buscan que no se confunda una salida automatizada con una comunicación humana o un registro auténtico. La transparencia tiene que ser pertinente para el uso: una etiqueta genérica puede no bastar si la persona necesita conocer el papel del sistema para ejercer un derecho o decidir cómo actuar.

El Reglamento establece además obligaciones para proveedores de modelos de IA de propósito general. Estas incluyen documentación, información a quienes integran el modelo en otros sistemas y, para determinados modelos que presentan riesgo sistémico, evaluaciones, mitigación de riesgos y notificación de incidentes, entre otros requisitos. La distinción entre modelo general y sistema de uso final es relevante:

un modelo puede ser incorporado a muchos productos y contextos, por lo que la gestión de riesgos se distribuye entre quienes desarrollan el modelo y quienes lo adaptan o despliegan.

La aplicación se organiza por etapas y ha recibido reformas posteriores. A la fecha de corte, EUR-Lex incorpora el Reglamento (UE) 2026/1744, que modifica el marco de 2024 para simplificar aspectos de la implementación de las reglas armonizadas. Por ello, las fechas y obligaciones de aplicación deben revisarse en el texto consolidado vigente y no repetirse mecánicamente a partir de versiones iniciales del *AI Act*. La existencia de una reforma no significa que el esquema de riesgos haya desaparecido; exige identificar qué disposición fue modificada y cuál es su nueva redacción.

5.1.3. Supervisión y alcances del modelo europeo

El Reglamento establece una arquitectura de gobernanza que combina autoridades nacionales, coordinación a nivel de la Unión y funciones de apoyo técnico. Su eficacia depende de que los organismos puedan acceder a información, evaluar riesgos, investigar posibles incumplimientos y coordinarse con las autoridades sectoriales. La supervisión no se limita a comprobar que exista documentación: requiere determinar si los sistemas funcionan dentro de su finalidad, si las medidas de control son adecuadas y si los derechos de las personas pueden hacerse valer.

El enfoque europeo ha sido descrito como basado en riesgos, pero debe interpretarse junto con sus límites. Una clasificación previa no puede anticipar todas las consecuencias de un sistema en cada contexto, y las categorías pueden requerir revisión cuando cambian los usos o la evidencia disponible. Además, el cumplimiento formal no sustituye la evaluación de daños concretos. Un sistema que cumple requisitos técnicos aún puede producir una decisión que deba revisarse conforme al derecho de igualdad, protección de datos, consumo o tutela judicial.

Para otras jurisdicciones, el *AI Act* puede servir como referencia comparada o adquirir relevancia práctica por el alcance de mercado de la Unión. No obstante, trasladar el modelo requiere valorar las instituciones y recursos disponibles en cada país. El enfoque basado en categorías de riesgo presupone definiciones, autoridades competentes, procedimientos de evaluación y capacidad de fiscalización. La copia legislativa sin una estructura institucional suficiente puede generar obligaciones difíciles de aplicar o dejar brechas entre la norma y la práctica.

5.2. Modelos regulatorios en Estados Unidos, China y otras jurisdicciones

5.2.1. Estados Unidos: regulación distribuida y orientación a la innovación

El modelo estadounidense se ha desarrollado mediante una combinación de leyes federales sectoriales, competencias de agencias, normas estatales, decisiones judiciales, estándares técnicos y directrices administrativas. A diferencia de la Unión Europea, no existe un único reglamento federal horizontal que reúna todas las obligaciones generales para sistemas de IA en los distintos sectores. La ausencia de un marco único no significa que la IA esté fuera de regulación: pueden aplicarse normas existentes sobre privacidad, consumo, empleo, salud, crédito, discriminación y seguridad.

La orientación de la política federal ha variado y, a la fecha de corte, se caracteriza por un énfasis fuerte en innovación, liderazgo tecnológico y coordinación nacional. La Orden Ejecutiva 14179, emitida en enero de 2025, estableció una política de eliminación de barreras a la innovación y revocó determinadas directrices anteriores. En diciembre de 2025, la Orden Ejecutiva 14365 impulsó la búsqueda de un marco nacional y planteó cuestionamientos a ciertas leyes estatales. El texto reconoce que una política nacional legislativa requeriría actuación del Congreso y ordena medidas ejecutivas dentro de las competencias existentes. Por

su naturaleza, una orden ejecutiva orienta y dirige al poder ejecutivo; no equivale a una ley general aprobada por el Congreso.

En marzo de 2026, la Casa Blanca presentó un marco legislativo nacional y señaló que requería colaboración con el Congreso para convertirse en ley. Mientras una propuesta no complete el procedimiento legislativo, debe describirse como iniciativa política o legislativa, no como norma vigente. Este punto es esencial en el análisis comparado: una estrategia, una orden ejecutiva, una guía de agencia y una ley federal tienen distinta jerarquía, alcance y permanencia.

Las leyes estatales pueden establecer requisitos propios sobre transparencia, toma de decisiones automatizadas, protección frente a discriminación, privacidad o uso de IA por organismos públicos. La coexistencia de reglas estatales y federales crea disputas sobre preeminencia, competencia y comercio interestatal. La orden ejecutiva de diciembre de 2025 instruyó a agencias federales a evaluar leyes estatales y a impugnar algunas cuando se consideren incompatibles con la política federal. Esta orientación ha generado tensión entre la meta de una regulación nacional uniforme y la capacidad de los estados para responder a riesgos locales.

El modelo también se apoya en estándares técnicos y marcos de gestión de riesgos. El NIST AI Risk Management Framework, de uso voluntario, ofrece funciones para identificar, medir y gestionar riesgos, pero no constituye por sí solo una ley. Las agencias pueden incorporar estándares o pautas en procesos de compra, regulación sectorial o gestión interna. De este modo, prácticas voluntarias pueden adquirir relevancia jurídica indirecta si un contrato, una norma o una decisión regulatoria las incorpora.

La experiencia estadounidense muestra una tensión entre agilidad y fragmentación. El enfoque distribuido puede facilitar respuestas específicas y permitir experimentar a nivel estatal, pero puede crear

incertidumbre para organizaciones que operan en varios estados. La búsqueda de una norma federal común pretende reducir esa dispersión; a la vez, desplazar reglas estatales puede disminuir la capacidad de adoptar protecciones más exigentes. La evaluación de este modelo requiere observar no solo el discurso de innovación, sino las obligaciones efectivamente promulgadas y los mecanismos disponibles para proteger a las personas.

5.2.2. China: gobernanza de servicios, contenidos y seguridad

China ha construido un marco de IA mediante leyes generales sobre ciberseguridad, seguridad de datos y protección de información personal, además de regulaciones y medidas dirigidas a tecnologías y servicios concretos. Entre estas se encuentran las reglas sobre recomendación algorítmica, síntesis profunda y servicios de IA generativa. Este enfoque integra el desarrollo tecnológico con objetivos de seguridad nacional, control del contenido, protección de intereses públicos e innovación industrial.

Las Medidas Provisionales para la Administración de Servicios de Inteligencia Artificial Generativa fueron promulgadas por la Administración del Ciberespacio de China junto con otros organismos y entraron en vigor el 15 de agosto de 2023. Se aplican a servicios que utilizan IA generativa para ofrecer al público dentro de China la generación de textos, imágenes, audio o video, conforme a su alcance. El texto distingue esos servicios públicos de la investigación, desarrollo o uso interno que no ofrece dichos servicios al público nacional.

Las medidas combinan promoción de la innovación con obligaciones sobre la legalidad de los datos de entrenamiento, la protección de información personal, la calidad de los datos, la transparencia, la exactitud y la fiabilidad de los contenidos. También establecen deberes de gestión de contenidos, canales de reclamación y medidas para responder ante usos ilícitos. Los proveedores pueden tener que cumplir

requisitos adicionales conforme a otras normas o a las características de la plataforma y del servicio.

Este modelo no se reduce a una sola ley integral. Opera a través de instrumentos administrativos, marcos generales y obligaciones técnicas o de contenido que se aplican a distintos tipos de servicios. Su lectura debe considerar el contexto constitucional e institucional chino y no asimilar las referencias a derechos, seguridad o interés público a las categorías jurídicas empleadas en otros ordenamientos. La comparación permite describir diferencias de objetivos regulatorios y mecanismos de aplicación, sin asumir que términos similares producen iguales efectos jurídicos.

5.2.3. Otras jurisdicciones: enfoques legislativos, sectoriales y voluntarios

Otros países adoptan estrategias distintas según sus prioridades y capacidades. Japón, por ejemplo, aprobó en 2025 una ley de promoción de la investigación, desarrollo y utilización de tecnologías relacionadas con IA. Su aproximación enfatiza el fomento, la inversión, la coordinación pública y la preparación institucional; opera junto con normas sectoriales y orientaciones existentes, en vez de reproducir sin más la estructura del *AI Act*. Las diferencias reflejan decisiones sobre qué instrumentos conviene utilizar y qué capacidades regulatorias existen.

Singapur ha desarrollado marcos de gobernanza y herramientas técnicas que promueven la evaluación y el uso responsable. Sus marcos de gobernanza para IA tradicional, generativa y, desde 2026, sistemas agenticos ofrecen prácticas para organizaciones, pruebas y controles. La autoridad de desarrollo digital describe algunas herramientas como guías voluntarias, útiles para implementar salvaguardias, pero distintas de una regulación vinculante de alcance general.

En América Latina, Brasil es un caso legislativo relevante por la tramitación del Proyecto de Ley 2338/2023, que propone un marco para el desarrollo y uso ético y responsable de la IA. A la fecha de corte, la ficha legislativa de la Cámara de Diputados registra el proyecto en trámite y pendiente de informe en comisión especial; por tanto, sus disposiciones no deben presentarse como derecho vigente. El proceso muestra una discusión sobre clasificación de riesgos, derechos de las personas, obligaciones de proveedores y usuarios y mecanismos institucionales.

La comparación internacional también debe incluir normas generales que regulan conductas sin mencionar expresamente la IA. Las leyes de protección de datos, derechos del consumidor, trabajo, no discriminación, propiedad intelectual y responsabilidad pueden aplicarse a sistemas automatizados si se cumplen sus condiciones. Un país sin ley integral de IA no está necesariamente desprovisto de garantías; el análisis debe identificar las normas sectoriales, las autoridades competentes y los vacíos específicos que persisten.

Los distintos modelos no son mutuamente excluyentes. Un país puede adoptar una estrategia nacional de innovación, establecer reglas sectoriales, aplicar una ley de datos personales, emitir directrices voluntarias y participar en acuerdos internacionales. Para comparar con rigor, conviene clasificar cada instrumento por su fuerza jurídica, alcance territorial, sujetos obligados, autoridad de supervisión, mecanismos de cumplimiento y recursos disponibles para las personas afectadas.

Tabla 5.1

Comparación de enfoques regulatorios seleccionados

Jurisdicción o marco	Tipo de enfoque	Características principales	Alcance y cautela para la comparación
Unión Europea	Regulación horizontal y basada en riesgos	El Reglamento de IA establece obligaciones diferenciadas según el riesgo, el tipo de sistema y la función de cada actor. Incluye prácticas prohibidas y requisitos para ciertos sistemas de alto riesgo.	Es vinculante dentro de su ámbito de aplicación. Debe leerse junto con las normas europeas de protección de datos, consumo, competencia y derechos fundamentales.
Estados Unidos	Regulación distribuida y sectorial	Combina normas aplicables a sectores y actividades, directrices y medidas de política pública.	No debe describirse como un único modelo normativo uniforme. Las obligaciones dependen de la norma, el sector y la jurisdicción pertinentes.
China	Regulación de servicios, contenidos y seguridad	Incluye reglas aplicables a servicios algorítmicos, contenidos sintéticos y servicios de IA generativa.	Las obligaciones dependen del servicio y de su contexto. No conviene reducir el enfoque a una sola norma ni equipararlo automáticamente al modelo europeo.

Jurisdicción o marco	Tipo de enfoque	Características principales	Alcance y cautela para la comparación
UNESCO	Orientación ética y de política pública	Su Recomendación propone principios centrados en los derechos humanos, la dignidad, la inclusión, la privacidad y la supervisión durante el ciclo de vida de la IA.	Es una recomendación internacional; por sí sola no equivale a una ley nacional ni a un tratado vinculante.
OCDE	Cooperación y principios intergubernamentales	Sus principios promueven una IA innovadora y confiable, el respeto de los derechos humanos, la transparencia, la robustez y la rendición de cuentas.	Es una recomendación. Orienta políticas y cooperación, pero no reemplaza las normas vinculantes de cada jurisdicción.
Consejo de Europa	Marco internacional de derechos humanos	La Convención Marco vincula las actividades relacionadas con la IA con los derechos humanos, la democracia y el Estado de derecho.	Su aplicación como tratado depende de la ratificación y entrada en vigor para cada Estado.

Nota. *Elaboración propia con base en el Reglamento (UE) 2024/1689, la Recomendación sobre la Ética de la IA de la UNESCO, la Recomendación de la OCDE sobre IA y la Convención Marco del Consejo de Europa. La tabla compara características generales; no sustituye el análisis de las normas aplicables a un caso concreto.*

5.3. Instrumentos de organismos internacionales: UNESCO, OCDE, ONU y Consejo de Europa

5.3.1. Recomendaciones, estándares y cooperación multilateral

La gobernanza internacional de la IA combina instrumentos no vinculantes, tratados, resoluciones, estándares técnicos y foros de cooperación. Esta pluralidad responde a que la tecnología atraviesa fronteras y sectores, mientras que la creación de obligaciones jurídicas globales requiere acuerdos entre Estados. Los instrumentos internacionales pueden establecer valores comunes, facilitar interoperabilidad y promover capacidades, aunque sus efectos dependen de la naturaleza del instrumento y de su implementación nacional.

La UNESCO adoptó en noviembre de 2021 la Recomendación sobre la Ética de la Inteligencia Artificial, aprobada por sus Estados miembros como primer estándar global de ética en esta materia. La Recomendación articula un enfoque centrado en derechos humanos, dignidad, equidad, transparencia, supervisión humana, sostenibilidad y proporcionalidad, y propone medidas para políticas públicas. Su condición es la de recomendación: orienta a los Estados y puede influir en leyes y políticas, pero no equivale a un tratado que imponga obligaciones directamente exigibles del mismo modo que un convenio ratificado.

La OCDE adoptó sus Principios de IA en 2019 y los revisó en 2024 para reflejar avances como la IA generativa y fortalecer la implementación. La recomendación promueve innovación confiable, respeto por derechos humanos y valores democráticos, transparencia, robustez, seguridad, rendición de cuentas y crecimiento inclusivo. La OCDE ofrece también definiciones y guías prácticas que facilitan la comparación entre políticas nacionales. Estos instrumentos son recomendaciones intergubernamentales; su adopción no reemplaza

por sí sola la legislación interna, aunque pueden informar estrategias, estándares y reformas.

En las Naciones Unidas, la Asamblea General aprobó en marzo de 2024 la Resolución A/RES/78/265 sobre el aprovechamiento de oportunidades de sistemas de IA seguros, protegidos y confiables para el desarrollo sostenible. La resolución reconoce oportunidades y desafíos, llama a una cooperación inclusiva y subraya la necesidad de reducir brechas digitales y apoyar capacidades en países en desarrollo. Como resolución de la Asamblea General, orienta y expresa compromisos políticos multilaterales; no constituye por sí misma una convención especializada de IA con obligaciones equivalentes a las de un tratado.

La cooperación internacional también se desarrolla mediante foros como el G7, el G20, la Alianza Global sobre IA y espacios regionales. Estos mecanismos facilitan intercambio de prácticas, pruebas de seguridad, estándares y coordinación de políticas. Su flexibilidad permite incorporar actores técnicos y de la sociedad civil, pero su naturaleza puede ser menos vinculante y su representación no siempre reflejar equitativamente las prioridades de todos los países. La participación de países de ingresos medios y bajos es importante para que los estándares no dependan exclusivamente de capacidades y prioridades de economías tecnológicamente dominantes.

5.3.2. Tratados y marcos regionales de derechos humanos

La Convención Marco del Consejo de Europa sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho fue adoptada en 2024 y abierta a la firma el 5 de septiembre de ese año. Es un tratado internacional que busca que las actividades del ciclo de vida de los sistemas sean compatibles con derechos humanos, democracia y Estado de derecho. Sus obligaciones alcanzan a los Estados que queden vinculados conforme a las reglas del tratado; la firma, por sí sola, debe distinguirse de la ratificación y de la entrada en vigor para cada parte.

La Convención establece principios y obligaciones generales, entre ellos protección de derechos humanos, transparencia y supervisión, responsabilidad, igualdad y no discriminación, privacidad y protección de datos, fiabilidad e innovación segura. También contempla evaluaciones de riesgos e impactos, mecanismos de supervisión y remedios conforme a las obligaciones internacionales de cada parte. Su enfoque es tecnológicamente neutral y orientado a la compatibilidad del ciclo de vida con los estándares de derechos, en lugar de replicar exactamente la clasificación y las obligaciones de mercado del Reglamento europeo.

El Consejo de Europa también desarrolla metodologías de implementación, como HUDERIA, para evaluar riesgos e impactos desde la perspectiva de los derechos humanos, la democracia y el Estado de derecho. La metodología es orientación no vinculante asociada a la aplicación práctica, mientras que la Convención es un tratado. Esta distinción ilustra cómo los instrumentos voluntarios pueden apoyar la implementación de obligaciones jurídicas sin tener el mismo estatus que ellas.

En América, los estándares interamericanos sobre derechos humanos, privacidad y libertad de expresión pueden aplicarse a prácticas de IA a través de las obligaciones generales de los Estados y de las normas nacionales. Los principios de privacidad de la OEA ofrecen un marco regional de referencia para protección de datos, pero no constituyen una ley única para los Estados americanos. La aplicación concreta depende de tratados ratificados, constituciones, leyes y decisiones de los órganos competentes en cada país.

La existencia de instrumentos internacionales no elimina la diversidad jurídica. El mismo estándar puede incorporarse mediante una ley, una política pública, una guía de autoridad o un contrato. También puede haber diferencias en el alcance de las obligaciones, el nivel de protección y las vías de reparación. Por eso, la investigación comparada

debe identificar la norma, el estado de adhesión, su jerarquía interna y las instituciones responsables, en lugar de atribuir a "la comunidad internacional" una regla uniforme que todavía no existe.

Figura 5.1.

Capas de la gobernanza internacional de IA



Nota: *Elaboración propia. Diferenciar los tratados y normas vinculantes de recomendaciones y estándares voluntarios.*

5.4. El efecto Bruselas, la armonización normativa y la competencia regulatoria

5.4.1. Influencia externa y circulación de estándares

El "efecto Bruselas" describe la influencia que puede ejercer la regulación de la Unión Europea fuera de sus fronteras cuando el acceso a su mercado es importante y las empresas prefieren aplicar un estándar común en varios territorios. Si adaptar un producto a normas europeas resulta menos costoso que mantener versiones diferenciadas, una empresa puede extender determinadas prácticas a nivel global. La influencia también puede ocurrir cuando otros países usan el Reglamento como modelo legislativo o cuando sus definiciones y estándares

técnicos se incorporan a cadenas de suministro internacionales.

En IA, esta influencia puede manifestarse mediante documentación, evaluaciones de riesgo, transparencia, gobernanza de datos, auditorías y requisitos de supervisión. Una entidad ubicada en América Latina puede verse afectada indirectamente si presta servicios a clientes europeos, integra componentes de proveedores sujetos al marco europeo o comercializa productos en la Unión. La obligación jurídica directa depende de los criterios territoriales y materiales del Reglamento; las exigencias contractuales o comerciales son una vía distinta de difusión y no convierten automáticamente la norma europea en ley nacional.

La influencia regulatoria puede promover interoperabilidad. Si distintas jurisdicciones adoptan definiciones, métricas y requisitos compatibles, las organizaciones pueden reducir costos de cumplimiento y facilitar auditorías transfronterizas. Una base común también permite que autoridades cooperen y comparen incidentes. Sin embargo, la armonización no debe confundirse con uniformidad completa: los países pueden necesitar proteger derechos conforme a sus constituciones, regular sectores específicos o atender contextos lingüísticos y sociales propios.

La circulación de estándares puede ser asimétrica. Los países que controlan mercados grandes, infraestructura y capital tecnológico tienen mayor capacidad para imponer requisitos de acceso. Los países importadores pueden adoptar estándares externos sin contar con recursos para fiscalizar su cumplimiento o adaptar las obligaciones a sus realidades. Una armonización justa requiere participación en la elaboración de normas, transferencia de capacidades y atención a las necesidades de países en desarrollo, incluyendo disponibilidad de evaluadores, peritos y autoridades técnicas.

5.4.2. Armonización, divergencias y competencia regulatoria

La competencia regulatoria ocurre cuando los Estados buscan atraer inversión, talento o infraestructura tecnológica mediante reglas más favorables, o cuando procuran posicionarse como referentes de seguridad y confianza. Una regulación más flexible puede reducir costos inmediatos para desarrolladores, pero también trasladar riesgos a personas, trabajadores y comunidades. Un marco más estricto puede aumentar protección y confianza, aunque si carece de proporcionalidad o apoyo a la innovación puede dificultar la entrada de empresas pequeñas o la investigación.

La armonización puede reducir incertidumbre si define obligaciones compatibles y evita que un proveedor deba reconstruir su gobernanza para cada país. Pero también puede convertirse en un estándar mínimo que debilite protecciones más amplias si se impone sin debate democrático. La cooperación debe establecer objetivos comunes y, a la vez, permitir que las jurisdicciones adopten salvaguardias adicionales cuando sus obligaciones constitucionales o internacionales lo requieran.

La convergencia técnica ofrece una vía práctica. Normas sobre gestión de riesgos, calidad de datos, registros, pruebas, seguridad y documentación pueden ayudar a establecer evidencias comparables entre mercados. Los estándares no son neutrales: su selección define qué se mide, qué se considera un riesgo aceptable y quién puede realizar una evaluación. La participación de autoridades, academia, sociedad civil, sectores productivos y grupos afectados mejora la legitimidad y reduce la posibilidad de que el estándar refleje solo los intereses de quienes lo desarrollan.

La competencia también se expresa en la oferta de modelos normativos. La Unión Europea presenta un marco horizontal basado en riesgos y derechos; Estados Unidos prioriza una arquitectura federal y sectorial con fuerte orientación a innovación, mientras persisten disputas con

reglas estatales; China combina promoción, control de servicios y objetivos de seguridad; Singapur promueve marcos técnicos y voluntarios; y Brasil debate un proyecto legislativo integral. Esta comparación no determina cuál modelo es superior en abstracto. Cada uno refleja instituciones, prioridades políticas y capacidades diferentes, y debe evaluarse por sus resultados, garantías y mecanismos de cumplimiento.

Para América Latina, la armonización puede facilitar operaciones transfronterizas y proteger derechos de personas que utilizan servicios globales. Pero la región necesita una voz propia en la construcción de estándares y capacidades. La adopción de modelos externos debe considerar pluralidad cultural y lingüística, concentración de proveedores, protección de datos, dependencia de infraestructura extranjera y recursos de las autoridades. El objetivo no es elegir entre innovación o derechos, sino diseñar reglas aplicables, proporcionales y fiscalizables que permitan innovar sin dejar sin protección a las personas afectadas.

El derecho comparado debe actualizarse periódicamente. Un proyecto aprobado en una cámara puede seguir pendiente; una orden ejecutiva puede cambiar con una nueva administración; una recomendación puede ser revisada; y un tratado puede entrar en vigor para ciertos Estados, pero no para otros. En consecuencia, todo análisis debe indicar fecha de corte, fuente oficial, estatus jurídico y jurisdicción. Esa disciplina evita describir como vigente una propuesta o atribuir alcance universal a un marco regional.

Capítulo 6.

Regulación de la inteligencia artificial en América Latina



INTRODUCCIÓN

América Latina y el Caribe no siguen una sola ruta frente a la inteligencia artificial (IA). Algunos países han adoptado estrategias nacionales; otros han aprobado leyes generales o reglamentos; varios debaten proyectos legislativos, y muchos todavía aplican a estos sistemas normas transversales sobre protección de datos, consumo, trabajo, contratación pública y derechos fundamentales. Por ello, para describir el panorama regional hay que distinguir con cuidado entre políticas públicas, normas vigentes y propuestas legislativas.

Este capítulo examina esa diversidad y sus implicaciones. Primero presenta las estrategias y los instrumentos normativos nacionales. Luego compara algunas experiencias y proyectos de ley representativos. A continuación, analiza cómo el sistema interamericano ofrece un marco de derechos aplicable a las decisiones automatizadas. Finalmente, aborda los obstáculos de implementación y las posibilidades de cooperación regional. El panorama normativo se refiere a la información disponible al **20 de octubre de 2026**; el estado de los proyectos legislativos puede cambiar.

6.1. Panorama normativo y estrategias nacionales de IA en la región

6.1.1. Un mapa desigual: políticas, leyes y reglas transversales

La gobernanza de la IA en América Latina se ha desarrollado mediante instrumentos de distinta naturaleza y fuerza jurídica. Una estrategia nacional expresa prioridades y orienta la acción del Estado, pero, por sí sola, no suele imponer obligaciones exigibles a empresas o administraciones. Una ley puede establecer derechos, prohibiciones, deberes, autoridades y sanciones. Un reglamento, concreta cómo se aplica una ley. A estos instrumentos se suman normas anteriores —como las leyes de protección de datos, consumo y trabajo— que pueden regular ciertos efectos de la IA, aunque no la mencionen de forma expresa.

Esta variedad importa. Un gobierno puede declarar que promoverá una IA confiable y centrada en las personas, pero esa declaración no garantiza que quien sufra una decisión automatizada tenga acceso a una explicación o a una vía efectiva para impugnarla. A la inversa, la falta de una ley general de IA no significa que exista un vacío jurídico completo: las decisiones automatizadas pueden afectar derechos ya protegidos por normas constitucionales, administrativas y sectoriales. La cuestión es si esos marcos ofrecen herramientas suficientes para prevenir, identificar y reparar daños que surgen de sistemas opacos, adaptativos o desplegados a gran escala.

Las estrategias nacionales suelen cubrir varios objetivos a la vez: investigación, formación de talento, infraestructura de datos y cómputo, adopción en el sector público y productivo, y principios éticos. El desafío es evitar que el componente de promoción quede desconectado de las salvaguardias. Si una estrategia financia la automatización de servicios públicos pero no define quién evalúa sus efectos, cómo se controla la calidad de los datos ni qué recursos tiene la persona afectada,

puede acelerar la adopción sin fortalecer la rendición de cuentas.

El *Índice Latinoamericano de Inteligencia Artificial* (ILIA) 2025, elaborado con participación del Centro Nacional de Inteligencia Artificial de Chile y la Comisión Económica para América Latina y el Caribe (CEPAL), profundiza en las capacidades y brechas de 19 países. Este tipo de medición ayuda a comparar condiciones habilitantes —como infraestructura, datos, talento y gobernanza—, aunque un índice no reemplaza el análisis jurídico: la preparación tecnológica de un país no equivale necesariamente a una protección adecuada de derechos.

Una segunda referencia regional es la Metodología de Evaluación del Estadio de Preparación para la IA (RAM) de la UNESCO. Esta metodología examina dimensiones institucionales, jurídicas, sociales, científicas y técnicas y busca apoyar la implementación de la Recomendación sobre la Ética de la IA, adoptada por la Conferencia General de la UNESCO en 2021. Su utilidad está en identificar brechas y capacidades antes de que se traduzcan en políticas concretas; no es una ley ni convierte automáticamente sus recomendaciones en obligaciones nacionales. La UNESCO registra procesos RAM en países latinoamericanos, entre ellos Argentina, Chile, Colombia, Paraguay, Perú y Uruguay, además de Ecuador y El Salvador.

Por tanto, no es adecuado presentar a la región como si ya contara con un régimen uniforme o como si todas las políticas públicas tuvieran el mismo valor legal. La comparación exige preguntar qué documento está vigente, qué autoridad lo ejecuta, qué presupuesto lo respalda y qué mecanismos permiten exigir su cumplimiento.

6.1.2. Estrategias nacionales y capacidad de dirección pública

Chile, Brasil y Uruguay ilustran la función de las estrategias nacionales, aunque sus trayectorias son distintas. Chile publicó su primera

Política Nacional de IA en 2021 y abrió un proceso de actualización en 2023–2024. La política organiza sus prioridades en tres ejes: factores habilitantes; desarrollo y adopción; y gobernanza y ética. El plan de acción informado por el Ministerio de Ciencia contempla 177 iniciativas coordinadas por 14 ministerios. Estos instrumentos ofrecen una hoja de ruta administrativa y política; no deben confundirse con una ley general de IA.

Brasil cuenta con la *Estratégia Brasileira de Inteligência Artificial* (EBIA), instituida mediante la Portaria MCTI n.º 4.617, de 6 de abril de 2021, y modificada por la Portaria n.º 4.979, de julio del mismo año. La estrategia organiza acciones en nueve ejes y procura impulsar investigación, innovación y uso responsable. En 2024, el país presentó además el *Plano Brasileiro de Inteligência Artificial 2024–2028*, de orientación programática. Una estrategia y un plan pueden coordinar inversión y prioridades; no sustituyen, sin embargo, la legislación que corresponde al Congreso ni resuelven por sí solos cuestiones como la responsabilidad por daños, la supervisión independiente o los recursos de las personas afectadas.

Uruguay aprobó su *Estrategia Nacional de Inteligencia Artificial 2024–2030* tras un proceso de consulta pública. La estrategia se propone impulsar un desarrollo y un uso ético, responsable y seguro, y articula sus líneas de acción en torno a gobernanza, capacidades y desarrollo de la IA. Se inserta en una trayectoria más amplia de gobierno digital y protección de datos. Su valor reside, en parte, en vincular la política de IA con capacidades estatales ya existentes; su eficacia dependerá de la continuidad institucional, la coordinación entre organismos y la traducción de sus objetivos en protocolos aplicables.

Argentina adoptó en 2023 recomendaciones para una inteligencia artificial fiable en el ámbito de la Administración Pública Nacional. El instrumento se orienta a quienes participan en el diseño, desarrollo e implementación de soluciones de IA en la administración. Este

tipo de guía puede influir en prácticas internas y elevar estándares de contratación y evaluación, pero tiene un alcance distinto al de una ley general de IA: no crea por sí misma un régimen integral de obligaciones para todos los actores públicos y privados.

Colombia avanzó con el Documento CONPES 4144 de 2025, que formaliza una política nacional para fortalecer capacidades de investigación, desarrollo, adopción y uso de la IA. El documento contiene acciones y asignaciones que orientan a distintas entidades estatales. El CONPES es un instrumento de política pública y planeación; no equivale a una ley aprobada por el Congreso. Esta distinción es relevante para determinar su alcance, su exigibilidad y la autoridad competente ante posibles infracciones.

En Ecuador, la UNESCO ha apoyado la evaluación RAM para identificar fortalezas y brechas del ecosistema nacional. Ese trabajo diagnóstico puede servir para articular una estrategia y orientar decisiones legislativas, pero sus recomendaciones no son normas directamente exigibles. La presencia de estas iniciativas muestra que la agenda regional no se limita a redactar leyes: también requiere instituciones capaces de medir riesgos, formar funcionarios, establecer prioridades y revisar programas de adopción.

Las estrategias presentan, al menos, tres ventajas. Permiten coordinar sectores que de otro modo trabajan por separado; facilitan inversiones y programas de formación; y hacen visibles prioridades de largo plazo que pueden orientar la regulación. Sus límites también son claros: pueden no contar con presupuesto estable; sus acciones pueden dispersarse entre entidades; y la participación social puede agotarse en la fase de consulta si no existe seguimiento público de los compromisos.

Una política nacional debería especificar quién responde por cada objetivo, qué recursos están disponibles, qué indicadores se utilizarán y cómo se publicarán los avances. También necesita reconocer que

la capacidad no se mide únicamente por el número de proyectos de IA desplegados. Un servicio puede automatizarse con rapidez y, al mismo tiempo, producir errores difíciles de detectar, excluir a personas que no cuentan con datos digitales o trasladar decisiones sensibles a proveedores privados. La calidad institucional incluye poder evaluar esas consecuencias y corregirlas.

6.1.3. Normas transversales y respuestas sectoriales

Incluso donde no hay una ley general de IA, las normas existentes pueden incidir directamente en los sistemas automatizados. Las leyes de protección de datos personales suelen establecer principios de finalidad, proporcionalidad, seguridad y derechos de acceso, rectificación o supresión; algunas jurisdicciones contemplan garantías frente a decisiones basadas exclusivamente en tratamiento automatizado. La eficacia de esas reglas depende de su cobertura, de las competencias de las autoridades de protección de datos y de que las personas puedan comprender cuándo un sistema influyó en una decisión que les afecta.

Las normas de consumo pueden aplicarse a publicidad personalizada, recomendaciones comerciales y servicios que utilizan modelos automatizados. La legislación laboral puede ser pertinente cuando algoritmos asignan tareas, evalúan desempeño o filtran postulaciones. El derecho administrativo, a su vez, impone límites a las decisiones públicas: competencia, motivación, transparencia, igualdad y posibilidad de impugnación. En salud, finanzas, educación y seguridad pública, las reglas sectoriales también pueden ser relevantes cuando la IA se incorpora a procesos existentes.

La dificultad práctica es la fragmentación. Una persona puede necesitar identificar qué organismo regula el producto, qué norma rige los datos utilizados, quién adoptó la decisión y qué autoridad recibe la reclamación. Las entidades públicas pueden, por su parte, carecer de equipos interdisciplinarios para examinar el sistema en conjunto. La coordinación entre autoridades —por ejemplo, entre protección de

datos, consumo, telecomunicaciones, trabajo y supervisión financiera— resulta decisiva para evitar tanto vacíos como órdenes contradictorias.

Las reglas generales necesitan adaptarse a los cambios técnicos sin perder determinación. Una ley que dependa de una definición demasiado estrecha puede quedar obsoleta cuando cambien las técnicas; una que defina la IA de manera excesivamente amplia puede someter herramientas rutinarias a cargas innecesarias. Una alternativa es combinar definiciones funcionales con obligaciones graduales, basadas en el contexto, el propósito y la gravedad del daño posible, junto con mecanismos periódicos de revisión.

Este diseño también debe ser realista sobre la diversidad productiva de la región. Muchos Estados y empresas adoptan sistemas desarrollados en otros países, mediante servicios en la nube o contratos con proveedores transnacionales. La autoridad local puede no tener acceso suficiente a los datos de entrenamiento, la documentación técnica o los registros de operación. Una política nacional que no considere esas dependencias corre el riesgo de imponer obligaciones formales que no puede supervisar, o de permitir que funciones esenciales del Estado dependan de servicios cuya arquitectura y condiciones se deciden en el extranjero.

En suma, el panorama regional se compone de estrategias, normas generales ya vigentes, regulación sectorial y propuestas legislativas en distintos estados de avance. Para evaluar la protección efectiva no basta con contar cuántos países publicaron una estrategia o anunciaron una ley. Hace falta analizar cómo se conectan esas piezas, quién las ejecuta y qué remedios están al alcance de las personas.

6.2. Proyectos de ley y experiencias regulatorias en países latinoamericanos

6.2.1. Perú: una ley aprobada y un reglamento vigente

Perú ofrece uno de los ejemplos más claros de legislación específica en la región. La Ley n.º 31814 promueve el uso de la IA en favor del desarrollo económico y social. En 2025 se aprobó su reglamento mediante el Decreto Supremo n.º 115-2025-PCM. Por tanto, al describir el caso peruano ya no corresponde hablar solamente de un proyecto de reglamento: la ley y su desarrollo reglamentario forman parte del marco vigente.

El diseño combina promoción y gobernanza, con particular relevancia para el uso de la IA en el sector público. La ley declara una finalidad de desarrollo, mientras que el reglamento concreta disposiciones para la gobernanza y el uso de sistemas. La implementación será una prueba crítica: deberá aclarar qué entidades supervisan, cómo se incorporan los riesgos a las decisiones de adquisición y despliegue y qué mecanismos permiten a las personas conocer e impugnar el uso de un sistema que incide en sus derechos.

El análisis jurídico no debe dar por sentado que la promulgación resuelve todos esos desafíos. Una ley puede crear un marco y dejar cuestiones importantes para su aplicación administrativa, para normas complementarias o para la interpretación judicial. También puede coexistir con regímenes de protección de datos y gobierno digital que deban coordinarse. La eficacia dependerá de la claridad de las responsabilidades, de la capacidad técnica de las entidades y de la transparencia de los sistemas utilizados por organismos públicos y privados.

6.2.2. Brasil: debate legislativo y revisión en la Cámara

Brasil ha mantenido un debate legislativo amplio. El Proyecto de Ley 2.338/2023 fue aprobado por el Senado en diciembre de 2024 y remitido a la Cámara de Diputados en marzo de 2025. En la Cámara se constituyó una comisión especial y el expediente registra actividad legislativa durante 2026. Por ello, al corte de este capítulo, la propuesta seguía en trámite y no debía presentarse como ley vigente.

El proyecto ha sido objeto de debate sobre un enfoque basado en riesgos, derechos y responsabilidades de los agentes que desarrollan, ofrecen o utilizan sistemas. Las versiones de un proyecto pueden cambiar durante su tramitación, de modo que cualquier análisis de sus obligaciones debe identificar el texto concreto y su estado parlamentario. La discusión brasileña refleja una tensión más amplia: cómo regular los usos con potencial de daño grave sin imponer cargas que frenen usos beneficiosos o que sean desproporcionadas para actores de menor tamaño.

La tramitación también pone de relieve la importancia de articular el régimen futuro con la protección de datos, el derecho del consumidor, la propiedad intelectual y las normas sectoriales. Una ley general puede ofrecer categorías y principios comunes, pero necesita reglas claras sobre supervisión, coordinación institucional y evaluación de impacto. Sin esas herramientas, el modelo corre el riesgo de consistir en una declaración de principios difícil de ejecutar.

Brasil también presenta experiencia en estrategias nacionales y planes de inversión pública. Esa coexistencia subraya que una ley general y una política de desarrollo cumplen funciones diferentes. La primera determina obligaciones y salvaguardias; la segunda moviliza infraestructura, investigación y capacidades. Una regulación eficaz necesita ambas dimensiones, pero debe preservar la independencia de la supervisión y la posibilidad de revisar proyectos estatales de alto riesgo.

6.2.3. Chile: proyecto legislativo después de la política nacional

Chile cuenta con una política nacional de IA y ha avanzado en una iniciativa legislativa. En octubre de 2025, la Cámara de Diputadas y Diputados aprobó el proyecto destinado a regular los usos de sistemas de IA y lo remitió al Senado. La fuente oficial consultada describe su avance a esa fecha; para valorar una norma como vigente se

requiere comprobar si el Congreso completó el trámite y si la ley fue promulgada. Al corte aquí utilizado, el proyecto debe tratarse como iniciativa legislativa, no confundirse con la política nacional de 2021 ni con su actualización.

La diferencia entre estos instrumentos ilustra una secuencia frecuente en la región: primero se adopta una política ejecutiva y luego se debate si es necesaria una ley, qué riesgos debe cubrir y qué autoridades deberán aplicarla. La política puede establecer principios y prioridades, pero no necesariamente habilita a una entidad para sancionar a un proveedor ni define por sí sola un derecho individual exigible.

La discusión chilena también muestra la relevancia de la consulta y la coordinación técnica. El Ministerio de Ciencia informó un proceso de participación para actualizar la política, con encuentros en distintas regiones y participación de actores académicos, sociales, empresariales y estatales. Esa experiencia puede enriquecer la legislación si las preocupaciones recogidas se traducen en disposiciones precisas: evaluación de impacto, transparencia, protección frente a discriminación y mecanismos de reclamación.

6.2.4. Colombia y Ecuador: política aprobada y propuestas en trámite

Colombia aprobó el CONPES 4144 de 2025 como política nacional de IA, mientras que en la Asamblea Nacional de Ecuador se han discutido proyectos de regulación y promoción de esta tecnología. La diferencia es sustantiva: el CONPES es un instrumento de política pública aprobado por el Consejo Nacional de Política Económica y Social; un proyecto presentado o calificado por una comisión legislativa aún no constituye ley. En Colombia, la implementación del CONPES orienta a las entidades y asigna acciones, pero no debe presentarse como una ley general de IA.

En Ecuador, el proyecto de Ley Orgánica de Regulación y Promoción de la Inteligencia Artificial fue presentado en junio de 2024 y recibió calificación del Consejo de Administración Legislativa en julio de ese año. La Asamblea informó posteriormente sobre el análisis y recepción de aportes para proyectos de IA. Esos registros permiten afirmar que hubo iniciativas legislativas y debate parlamentario; no autorizan a describir el proyecto como derecho vigente sin confirmar la aprobación final, sanción y publicación correspondiente.

La existencia de múltiples proyectos en un mismo país puede tener varias causas: distintas prioridades políticas, iniciativas presentadas por legisladores diferentes y debates sobre si el problema requiere una ley integral o ajustes a normas existentes. Para comprender el proceso legislativo no basta con citar el título de una propuesta. Debe verificarse su expediente, sus informes, la comisión responsable, las votaciones y el texto actualizado. En un campo en evolución, mantener esa trazabilidad es parte del rigor jurídico.

6.2.5. Patrones comparados y cautelas jurídicas

La comparación regional revela al menos cuatro patrones. Primero, estrategias ejecutivas que dirigen la investigación y la adopción sin establecer un régimen completo de responsabilidad. Segundo, propuestas legislativas que buscan una ley general basada en riesgos y derechos. Tercero, leyes que promueven la IA y desarrollan reglas para su uso, como el caso peruano. Cuarto, respuestas sectoriales que regulan problemas concretos —por ejemplo, decisiones administrativas automatizadas, publicidad o biometría— sin adoptar un marco único.

Ninguno de esos patrones debe evaluarse únicamente por su novedad. Una ley puede contener buenas categorías, pero carecer de autoridades dotadas de recursos, procedimientos de auditoría y vías de reclamación. Una estrategia puede no ser vinculante, pero producir cambios concretos si coordina entidades, fija presupuestos y publica resultados. La pregunta relevante es cómo se distribuyen las obligaciones a lo largo

del ciclo de vida: diseño, entrenamiento, prueba, compra, despliegue, monitoreo y retiro.

También conviene evitar dos simplificaciones. La primera es asumir que una ley regional será efectiva solo por inspirarse en la regulación de otra jurisdicción. La segunda es concluir que, en ausencia de una ley general, todo uso de IA queda fuera del derecho. En realidad, las reglas existentes siguen rigiendo y pueden ser exigibles, aunque a veces sean difíciles de aplicar a decisiones opacas o a cadenas de suministro complejas.

La importación de modelos regulatorios requiere considerar capacidades locales, estructura de mercado y garantías constitucionales. Un sistema de clasificación de riesgos, por ejemplo, puede ser útil si la autoridad puede examinar la documentación y exigir correcciones. Si no puede hacerlo, la clasificación puede depender principalmente de autodeclaraciones empresariales. De igual modo, una obligación de explicar una decisión debe considerar qué información necesita la persona para cuestionar el resultado y qué documentación debe conservar la institución.

La regulación también debe evitar que el costo de cumplimiento recaiga de forma desproporcionada en universidades, pequeñas empresas u organizaciones comunitarias mientras los grandes proveedores mantienen el control técnico y contractual. Esto no justifica reducir garantías, sino diseñar obligaciones proporcionales y fortalecer la capacidad pública de inspección. La cooperación entre autoridades puede ser más eficaz que multiplicar formularios o crear organismos sin presupuesto ni acceso a conocimiento especializado.

Tabla 6.1

Estado de las respuestas normativas y de política pública en países seleccionados

País	Instrumento o respuesta	Estado al 21 de octubre de 2026	Naturaleza y cautela jurídica
Perú	Ley n.º 31814, Decreto Supremo n.º 115-2025-PCM y Estrategia Nacional de Inteligencia Artificial 2026–2030	La ley cuenta con reglamento aprobado. También se aprobó la estrategia nacional para 2026–2030.	Marco legal y de política pública vigente. La estrategia orienta acciones; no debe confundirse con una ley.
Brasil	Proyecto de Ley n.º 2.338/2023	Continúa en trámite en la Cámara de Diputados y se encuentra vinculado a la comisión especial correspondiente.	Es una propuesta legislativa, no una ley vigente. Indicar su estado parlamentario a la fecha de cierre del libro.
Chile	Proyectos de ley sobre regulación y límites al desarrollo de la IA	Los proyectos identificados en el capítulo seguían en trámite legislativo, con informes y discusión en comisión durante 2025 y 2026.	Son iniciativas en debate; no presentarlas como normativa aprobada.
Colombia	Documento CONPES 4144 de 2025, Política Nacional de Inteligencia Artificial	Política aprobada con acciones previstas hasta 2030.	Es un instrumento de política pública; no equivale por sí mismo a una ley general de IA.

País	Instrumento o respuesta	Estado al 21 de octubre de 2026	Naturaleza y cautela jurídica
Ecuador	Proyectos de regulación y gobernanza de la IA	El proyecto de Regulación y Promoción de IA fue archivado en septiembre de 2026. En septiembre también se registró una nueva iniciativa sobre desarrollo, uso ético y gobernanza de IA, cuyo trámite debe diferenciarse del proyecto archivado.	No afirmar que existe una ley general de IA vigente. Verificar el estado de la nueva propuesta antes del cierre editorial.

Nota. *Elaboración propia con base en fuentes oficiales consultadas al 21 de octubre de 2026. El estado legislativo es temporal y puede cambiar; actualizar esta tabla antes de publicar el libro.*

6.3. Sistema interamericano de derechos humanos y estándares aplicables a la IA

6.3.1. Un marco tecnológicamente neutral

El sistema interamericano no necesita una convención dedicada exclusivamente a la IA para que sus garantías sean aplicables a los efectos de decisiones automatizadas. Los tratados de derechos humanos protegen a las personas frente a actos estatales; sus obligaciones pueden abarcar tecnologías utilizadas por autoridades públicas y, conforme a las reglas de responsabilidad internacional, ciertas omisiones estatales ante daños causados por actores privados. La tecnología empleada no elimina la obligación de respetar y garantizar los derechos.

La Convención Americana sobre Derechos Humanos (CADH) reconoce, entre otros, el derecho a la integridad personal, la libertad personal, las garantías judiciales, la vida privada, la libertad de pensamiento y expresión, la igualdad ante la ley y la protección judicial.

Estos derechos son pertinentes cuando un sistema automatizado influye en la asignación de beneficios, la vigilancia, la selección laboral, la evaluación educativa, la investigación penal o la moderación de contenidos.

La consecuencia práctica es que el examen jurídico debe centrarse en el efecto de la decisión y en la participación del Estado, no solamente en la etiqueta técnica del sistema. Si una autoridad utiliza una puntuación algorítmica para priorizar inspecciones o negar un servicio, debe poder justificar la legalidad y finalidad de la medida, proteger a las personas frente a errores y ofrecer un procedimiento para controvertirla. Un proveedor privado no desplaza la responsabilidad de la entidad pública que decide usar la herramienta para ejercer una función estatal.

El enfoque interamericano exige además valorar el contexto y los impactos diferenciados. Un sistema que parece neutral puede reproducir patrones de discriminación si los datos contienen desigualdades históricas o si las variables utilizadas sirven como sustitutos de características protegidas. El análisis debe mirar no solo la intención del diseñador, sino también quién soporta los errores, a quién se excluye, qué grupos son objeto de vigilancia y qué posibilidades reales tienen de reclamar.

6.3.2. Privacidad, datos personales y vigilancia

El artículo 11 de la CADH protege la honra, la dignidad y la vida privada frente a injerencias arbitrarias o abusivas. La recolección masiva de datos, el reconocimiento facial, la inferencia de atributos sensibles y la combinación de bases de datos pueden afectar ese ámbito incluso cuando los datos individuales parecen públicos o se obtienen de fuentes distintas. La cuestión jurídica no se agota en si una persona aceptó una política de privacidad: también importan la finalidad, la necesidad, la proporcionalidad, la seguridad y las posibilidades de oposición o corrección.

En el sector público, el uso de IA para vigilancia plantea riesgos específicos. La identificación biométrica en espacios públicos, la predicción de conductas o la selección automatizada de individuos pueden aumentar la intensidad de la intervención estatal y afectar la libertad de reunión, expresión y circulación. La legalidad exige una base normativa clara y controles adecuados; la necesidad requiere justificar por qué el objetivo no puede alcanzarse con medios menos intrusivos; y la proporcionalidad obliga a ponderar los beneficios alegados y el daño potencial. Estos parámetros no se satisfacen con una referencia genérica a la seguridad pública.

La falta de transparencia puede volver ilusorios los derechos de acceso y rectificación. Una persona puede ignorar que su perfil, puntuación o historial fue procesado y, por ello, no saber qué pedir ni qué decisión impugnar. Las entidades que despliegan estos sistemas necesitan registrar su finalidad, origen de datos, categorías de personas afectadas, proveedor, pruebas realizadas y mecanismos de revisión. Esa documentación no es un fin burocrático: hace posible que la autoridad y la persona afectada examinen la decisión.

El acceso a una explicación debe entenderse de forma funcional. No siempre será necesario revelar código fuente ni todos los parámetros técnicos. Sí puede ser indispensable comunicar qué datos relevantes se utilizaron, qué factores influyeron en el resultado, qué entidad tomó la decisión, cuál es el grado de intervención automatizada y cómo solicitar revisión humana. La explicación debe permitir cuestionar errores, sesgos o aplicación de criterios ilegales, y no reducirse a una fórmula genérica.

6.3.3. Igualdad, debido proceso y acceso a remedios

La igualdad y la no discriminación, recogidas en los artículos 1.1 y 24 de la CADH, son centrales cuando un algoritmo clasifica a personas o distribuye oportunidades. Un modelo puede producir resultados diferentes entre grupos por variables como territorio, idioma, historial

de interacción o nivel de ingresos, aunque no utilice explícitamente categorías protegidas. La evaluación debe examinar los datos, los resultados por grupo, el contexto de uso y la posibilidad de que el sistema refuerce exclusiones existentes.

El debido proceso es relevante cuando una decisión automatizada afecta derechos u obligaciones. El artículo 8 protege garantías judiciales en la determinación de derechos y obligaciones de orden civil, laboral, fiscal o de cualquier otro carácter. La Convención no contiene una regla técnica única para todo sistema de IA, pero sí exige garantías adecuadas al procedimiento concreto. En una decisión administrativa automatizada, esto puede requerir que la autoridad informe sobre el uso del sistema, permita conocer los elementos esenciales del expediente, motive la decisión y ofrezca revisión por una persona competente.

La revisión humana no debe convertirse en una firma formal sobre un resultado que nadie puede cuestionar. Para que sea efectiva, quien revisa necesita autoridad para apartarse de la recomendación automatizada, acceso a información suficiente, capacitación y tiempo. También deben existir reglas que eviten que el funcionario confíe de forma acrítica en el resultado por su apariencia técnica o por la presión institucional para automatizar.

El artículo 25 de la CADH exige un recurso efectivo ante violaciones de derechos. Aplicado a decisiones automatizadas, ello plantea preguntas prácticas: ¿puede la persona saber que un algoritmo participó?, ¿qué autoridad tiene competencia?, ¿es posible acceder a las pruebas relevantes?, ¿se conserva el registro del sistema?, ¿puede suspenderse provisionalmente una decisión potencialmente lesiva? Si las respuestas son negativas, el remedio puede existir formalmente y fallar en la práctica.

La responsabilidad no puede quedar diluida entre múltiples participantes. Una administración puede adquirir un sistema a un

proveedor, que a su vez utiliza modelos de terceros y conjuntos de datos externos. Sin embargo, frente a la persona afectada, la entidad pública debe identificar quién decidió utilizar el sistema y quién puede revisar la medida. Los contratos deberían exigir documentación, acceso a registros, cooperación con auditorías y conservación de evidencia suficiente para atender reclamaciones.

6.3.4. Libertad de expresión, democracia y estándares recientes

La libertad de expresión del artículo 13 de la CADH comprende buscar, recibir y difundir información e ideas. Los sistemas algorítmicos pueden incidir en ella mediante la clasificación y recomendación de contenidos, la moderación automatizada, la generación de información sintética o la vigilancia de periodistas y activistas. Las restricciones estatales, incluidas las que se implementan a través de plataformas o proveedores, deben ajustarse a las condiciones de legalidad, finalidad legítima, necesidad y proporcionalidad.

La Declaración de Principios sobre Libertad de Expresión de la CIDH, aunque no es un tratado, ofrece criterios interpretativos relevantes. Entre ellos figuran el acceso a la información, la protección frente a presiones directas o indirectas y la preservación del pluralismo. En entornos mediados por IA, estos principios invitan a preguntar si las decisiones automatizadas reducen de manera opaca la visibilidad de ciertos discursos, si amplifican campañas de manipulación o si las medidas de moderación permiten impugnación y reparación.

La CIDH y otras relatorías regionales han actualizado su trabajo frente a los efectos de la IA en el ecosistema informativo. En 2025, una declaración conjunta de mandatos sobre libertad de expresión abordó riesgos relacionados con IA, medios y autonomía; en 2026, la Relatoría Especial para la Libertad de Expresión abrió una consulta para elaborar un informe sobre el impacto de la IA —incluida la generativa— en los derechos humanos en las Américas. Estos documentos no son leyes

vinculantes, pero constituyen desarrollos interpretativos y de política pública que conviene seguir.

La cuestión democrática no se limita a la desinformación. También incluye la concentración de infraestructura y capacidad de decisión, la protección de datos de votantes, el uso de sistemas de perfilamiento político y la posibilidad de que contenidos sintéticos afecten el debate público. La regulación debe evitar tanto la inacción ante daños previsibles como la censura previa amplia o las obligaciones ambiguas que podrían utilizarse para restringir expresiones legítimas.

En seguridad y vigilancia digital, el informe temático publicado por la Relatoría Especial para la Libertad de Expresión de la CIDH en 2026 ofrece una referencia regional para evaluar tecnologías que pueden facilitar la persecución de periodistas y otras personas. Aunque su objeto no es exclusivamente la IA, sus preocupaciones resultan pertinentes cuando modelos automatizados se integran en sistemas de vigilancia, identificación o análisis de redes.

Por último, los estándares interamericanos deben aplicarse con atención a grupos históricamente discriminados y a comunidades en situación de vulnerabilidad. La obligación de igualdad sustantiva exige examinar impactos sobre pueblos indígenas, personas afrodescendientes, mujeres, personas con discapacidad, niños y niñas, personas migrantes y quienes viven en zonas con menor conectividad. La neutralidad de una herramienta no se presume por la ausencia de una categoría explícita: debe probarse considerando sus resultados y sus condiciones de uso.

Figura 6.1.

De los estándares interamericanos a la implementación nacional



Nota: *Elaboración propia. Evitar sugerir que un estándar internacional se aplica de idéntica forma en todos los ordenamientos.*

6.4. Desafíos de implementación, capacidades institucionales y cooperación regional

6.4.1. De los principios a la supervisión efectiva

El primer desafío regional es convertir principios en rutinas institucionales. Una ley o estrategia necesita autoridades con competencias delimitadas, personal interdisciplinario, acceso a información, presupuesto y procedimientos para actuar. La supervisión de sistemas de IA requiere conocimientos jurídicos, estadísticos, informáticos y sectoriales. Si una entidad carece de personal que pueda interpretar documentación técnica o revisar impactos, la norma puede quedar en manos de los propios proveedores.

La evaluación de impacto debe realizarse antes del despliegue en usos de riesgo elevado y actualizarse cuando cambien el propósito, los datos o las condiciones de operación. No basta con aplicar una lista de verificación al momento de comprar el sistema. La institución debe probarlo con poblaciones y contextos pertinentes, documentar limitaciones conocidas, establecer métricas para detectar errores y revisar los resultados una vez que el sistema está en funcionamiento.

La contratación pública es una palanca fundamental. Los Estados pueden exigir a sus proveedores documentación sobre datos, pruebas, desempeño, seguridad y límites de uso; conservar registros auditables; establecer derechos de inspección y cooperación; y prohibir reutilizaciones incompatibles con el propósito inicial. Sin embargo, las condiciones de compra deben reconocer las asimetrías de negociación y evitar aceptar cláusulas de secreto comercial que impidan explicar decisiones públicas o investigar posibles discriminaciones.

La rendición de cuentas necesita información pública accesible. Un registro de sistemas utilizados por organismos estatales puede incluir la entidad responsable, la finalidad, las poblaciones afectadas, el nivel de intervención automatizada y los mecanismos de reclamo, protegiendo los datos personales y los secretos legítimos. Publicar esta información permite que la ciudadanía, la academia y los órganos de control identifiquen usos que requieren escrutinio reforzado.

6.4.2. Brechas de infraestructura, talento y representación

La región enfrenta diferencias importantes en capacidad de cómputo, disponibilidad de datos de calidad, inversión en investigación y formación especializada. Esas brechas condicionan la autonomía de los Estados para desarrollar modelos propios y también su capacidad para revisar los sistemas que importan. La dependencia de proveedores externos puede reducir costos y acelerar servicios, pero plantea preguntas sobre alojamiento de datos, continuidad del servicio, transferencia internacional, auditabilidad y dependencia tecnológica.

La soberanía tecnológica no debería convertirse en una consigna para justificar sistemas menos respetuosos de derechos o restricciones a la circulación de conocimiento. Debe entenderse como la capacidad pública y social para elegir, evaluar, adaptar o rechazar tecnologías, y para sostener infraestructura y conocimiento compatibles con necesidades locales. Ello requiere cooperación con universidades, centros de investigación, empresas y organizaciones sociales, además

de políticas de formación que no se limiten a perfiles técnicos.

La representación lingüística y cultural también es una cuestión de calidad jurídica y técnica. Sistemas entrenados principalmente con datos de otras regiones pueden funcionar mal en español, portugués, lenguas indígenas y variedades locales; pueden desconocer instituciones, nombres, prácticas y formas de comunicación propias. El daño no es solo una menor precisión: una herramienta inadecuada puede restringir acceso a servicios o producir errores sistemáticos en grupos que ya enfrentan barreras institucionales.

La participación pública debe ocurrir antes y durante el ciclo de vida del sistema. Consultar al final, cuando la herramienta ya fue adquirida y desplegada, limita la posibilidad de revisar alternativas. Las comunidades afectadas y los trabajadores que interactúan con el sistema suelen identificar riesgos que no aparecen en la evaluación técnica. Su participación debe influir en la decisión, con mecanismos para responder a sus observaciones y documentar cambios.

6.4.3. Cooperación regional y estándares compartidos

La cooperación regional puede ayudar a reducir costos y evitar que cada país construya desde cero sus herramientas de evaluación. Los Estados podrían compartir modelos de auditoría, protocolos de pruebas, cláusulas de contratación pública, métodos de evaluación de sesgos y mecanismos de formación de funcionarios. Las metodologías de la UNESCO y los indicadores del ILIA proporcionan insumos para esta labor, aunque deben adaptarse a los marcos constitucionales y prioridades de cada país.

La cooperación también puede fortalecer la supervisión de servicios transfronterizos. Si un proveedor ofrece un sistema a varios países, las autoridades nacionales pueden intercambiar información sobre incidentes, vulnerabilidades y medidas correctivas. Los acuerdos deben

proteger la confidencialidad y los datos personales, pero evitar que la fragmentación jurisdiccional impida investigar fallas que afectan a varias poblaciones.

En derechos humanos, el sistema interamericano ofrece un marco común que puede orientar la convergencia sin exigir uniformidad legislativa. Los Estados pueden coordinar criterios sobre transparencia en decisiones públicas, evaluación de impactos diferenciados, límites a la vigilancia biométrica y vías de reparación. La convergencia más valiosa no consiste necesariamente en copiar el mismo texto legal, sino en acordar garantías mínimas y asegurar que las personas puedan reclamar de forma efectiva.

La cooperación debe incluir a la sociedad civil, academia, sindicatos, empresas y comunidades afectadas. La regulación de IA no puede apoyarse únicamente en expertos técnicos o en organismos económicos, porque las decisiones sobre aceptabilidad del riesgo, distribución de beneficios y límites a la vigilancia son también decisiones políticas y jurídicas. Los procesos de consulta deben ofrecer información comprensible, plazos razonables y respuestas motivadas a las propuestas recibidas.

6.4.4. Prioridades para una implementación sostenible

Una agenda regional de implementación puede organizarse alrededor de cinco prioridades. La primera es construir inventarios de sistemas públicos y evaluar con anticipación aquellos que inciden en derechos o servicios esenciales. La segunda es fortalecer a las autoridades existentes —protección de datos, consumo, trabajo, competencia, justicia y supervisión sectorial— antes de crear nuevas instituciones sin atribuciones claras. La tercera es garantizar revisión humana significativa y vías de reclamación para las personas afectadas.

La cuarta prioridad es invertir en capacidades: alfabetización pública,

formación de funcionarios, investigación local, documentación en español y portugués y apoyo a comunidades lingüísticas diversas. La quinta es sostener cooperación regional para compartir conocimientos, recursos y alertas, sin perder la capacidad de cada Estado para decidir qué usos son compatibles con sus obligaciones y su contexto.

Estas prioridades deberían traducirse en metas verificables. Por ejemplo: publicar los sistemas de alto impacto utilizados por entidades estatales; exigir evaluaciones de impacto antes de desplegarlos; medir diferencias de error entre grupos relevantes; conservar registros suficientes para investigar decisiones; y establecer plazos de respuesta a reclamaciones. Los indicadores deben medir protección y calidad, no solo cantidad de sistemas adoptados.

La evaluación periódica es esencial. Los modelos, los datos y los usos cambian; una herramienta aprobada para una función puede terminar utilizada para otra. La supervisión debe contemplar incidentes, cambios de proveedor, actualizaciones del modelo y nuevas poblaciones afectadas. También debe existir la posibilidad de suspender o retirar un sistema cuando el daño no pueda mitigarse, cuando la base legal sea insuficiente o cuando su desempeño no alcance los requisitos previstos.

La innovación y las garantías no son objetivos incompatibles, pero sí requieren decisiones explícitas sobre quién asume los costos y quién obtiene los beneficios. Una política que mide el éxito exclusivamente por productividad puede ignorar exclusiones, pérdida de privacidad o errores que recaen sobre las personas con menos capacidad de impugnar. Una política centrada en derechos debe observar los resultados concretos y ofrecer herramientas para corregirlos.



Capítulo 7.

Responsabilidad jurídica por daños causados por sistemas de IA



INTRODUCCIÓN

La inteligencia artificial puede contribuir a causar daños sin ser, por sí misma, sujeto de derechos u obligaciones. La respuesta jurídica debe identificar a las personas y organizaciones que diseñan, ofrecen, integran, despliegan o utilizan el sistema, así como la relación entre sus decisiones y el perjuicio. Esa tarea se complica cuando el resultado depende de múltiples participantes, cambios posteriores al despliegue y procesos técnicos que la persona afectada no puede conocer.

Este capítulo analiza los principales regímenes de responsabilidad aplicables: la responsabilidad civil por culpa o riesgo y la responsabilidad por producto defectuoso; la atribución entre actores de la cadena; las respuestas penales y administrativas; y las dificultades de prueba, seguro y reparación. El análisis distingue entre reglas vigentes y propuestas legislativas, y evita tratar la "autonomía" de un sistema como una explicación que, por sí sola, exima a quienes tenían deberes de cuidado, control o supervisión.

7.1. Responsabilidad civil: culpa, riesgo y responsabilidad por producto defectuoso

7.1.1. Funciones y elementos de la responsabilidad civil

La responsabilidad civil cumple, de manera general, dos funciones relacionadas: reparar el daño sufrido y orientar la conducta futura mediante incentivos de prevención. Cuando una aplicación automatizada genera una lesión física, una pérdida económica, una afectación reputacional o una exclusión injustificada, la primera pregunta jurídica no es si el sistema "decidió" por sí mismo. Es si concurren los elementos que el ordenamiento aplicable exige para atribuir responsabilidad: un daño jurídicamente relevante, una conducta o actividad vinculada con el demandado, un criterio de imputación y una relación causal suficiente.

La terminología varía entre jurisdicciones. Se habla de responsabilidad extracontractual, aquiliana o por hecho ilícito; en otros contextos, de responsabilidad contractual, de consumo, por riesgo o por producto. Esa diferencia no es puramente nominal: determina qué deber debe probarse, quién puede demandar, qué daños son indemnizables, qué plazos se aplican y cómo se distribuye la carga probatoria.

En un régimen basado en culpa, la víctima debe acreditar que el demandado incumplió un deber de cuidado. En sistemas que reconocen responsabilidad objetiva para ciertos supuestos, puede bastar la demostración del daño, de la actividad o producto relevante y del nexo causal, sin probar negligencia subjetiva. La responsabilidad objetiva no significa responsabilidad automática: suelen existir límites causales, defensas y requisitos legales. Tampoco toda actividad que incorpora IA es, por ese solo hecho, una actividad de riesgo jurídicamente calificada.

La culpa debe analizarse de acuerdo con el papel del actor y el contexto de uso. Para un proveedor, puede consistir en diseñar o comercializar un sistema sin pruebas adecuadas, no informar limitaciones conocidas

o mantener una herramienta pese a incidentes previsibles. Para la entidad que lo despliega, puede residir en utilizarlo para una finalidad incompatible con su diseño, prescindir de revisión humana cuando era necesaria o no establecer mecanismos de monitoreo. Para un usuario profesional, puede consistir en aceptar un resultado sin ejercer la revisión que exigían sus obligaciones. Los deberes concretos dependen del sector, la información disponible y la normativa aplicable.

La previsibilidad también requiere una valoración cuidadosa. Que un sistema produzca un resultado inesperado no demuestra automáticamente que fuera imprevisible para todos los actores. Debe examinarse qué pruebas se realizaron, qué advertencias se emitieron, qué fallas se conocían, qué cambios se hicieron después del despliegue y qué capacidad tenía cada participante para impedir o reducir el daño. Matthias (2004) denominó "brecha de responsabilidad" al problema que surge cuando los sistemas de aprendizaje producen conductas difíciles de anticipar con los métodos de control tradicionales. El concepto ayuda a describir una dificultad de atribución, pero no demuestra que exista un vacío legal inevitable ni que la máquina deba ser considerada responsable.

7.1.2. Culpa, riesgo y deber de cuidado

La aplicación de la culpa ordinaria a sistemas de IA exige traducir el deber de cuidado en prácticas verificables. En el desarrollo, eso puede incluir pruebas razonables según el propósito, documentación de limitaciones, protección frente a usos previsiblemente indebidos y evaluación de riesgos conocidos. En el despliegue, puede requerir determinar si el sistema es adecuado para la población y el entorno reales, capacitar a quienes lo utilizan y definir cuándo debe intervenir una persona. Después de la puesta en servicio, el deber puede comprender vigilancia de incidentes, actualización segura y suspensión cuando surgen fallas graves.

La diligencia no se determina exclusivamente por el resultado.

Un sistema puede causar un daño aun cuando se hayan adoptado precauciones razonables, y un daño grave puede revelar que los controles eran insuficientes. El análisis debe apoyarse en evidencia técnica y organizacional, sin convertir un estándar voluntario o una certificación en prueba concluyente de cumplimiento jurídico. Las normas técnicas y protocolos de gestión pueden servir como elementos de evaluación, pero el juez o la autoridad debe considerar el deber legal aplicable, el propósito del sistema y el contexto concreto.

La responsabilidad por riesgo plantea otra cuestión: ¿quién debe soportar el costo de los daños asociados a una actividad que genera beneficios y controla los peligros? Algunos ordenamientos atribuyen responsabilidad objetiva a actividades o productos determinados, a menudo para no exigir a la víctima una prueba casi imposible sobre la conducta interna del productor. La incorporación de IA no transforma automáticamente una actividad en "riesgosa" en sentido jurídico. La calificación debe provenir de la legislación o de una interpretación judicial compatible con las reglas de cada jurisdicción.

La idea de riesgo también puede orientar reformas legislativas. Si el sistema opera en un campo de alto impacto, el actor que obtiene beneficios, decide el despliegue o controla la información puede estar en mejor posición para prevenir y asegurar los daños que la persona afectada. Pero diseñar una regla de responsabilidad requiere equilibrar reparación, prevención, acceso a la innovación y protección frente a cargas desproporcionadas. El riesgo no debe convertirse en una fórmula vaga que haga responsable a cualquier participante sin atender a su conducta, control ni relación causal.

7.1.3. Producto defectuoso y software con IA

La responsabilidad por producto defectuoso resulta pertinente cuando un sistema de IA se incorpora a un producto o servicio que causa daño por una falla de seguridad. En términos generales, el examen pregunta si el producto ofrecía la seguridad que legítimamente podía esperarse,

considerando su presentación, uso razonablemente previsible y momento de puesta en circulación. La dificultad está en determinar quién es el productor cuando software, modelo, datos, componente físico y actualizaciones provienen de entidades distintas; y cuándo un defecto aparece en el diseño, en la información proporcionada o en el mantenimiento.

La Unión Europea aprobó la Directiva (UE) 2024/2853 sobre responsabilidad por productos defectuosos, que incorpora software y sistemas de IA a la definición de producto y contempla supuestos vinculados a actualizaciones y control del productor. Los Estados miembros deben transponerla antes del 9 de diciembre de 2026; la propia Directiva excluye de su ámbito los productos introducidos en el mercado o puestos en servicio antes de esa fecha. Por tanto, al corte de este capítulo, la directiva ya está aprobada, pero el plazo de transposición aún no ha vencido y su aplicación depende de las reglas temporales y de implementación correspondientes.

El nuevo régimen europeo puede ser útil para el análisis comparado, pero no se aplica automáticamente en países latinoamericanos. En América Latina, la respuesta debe buscarse en las normas nacionales de consumo, productos, daños y contratos. En Brasil, por ejemplo, el Código de Defesa do Consumidor prevé responsabilidad independiente de culpa para ciertos daños causados por defectos de productos y servicios, con reglas propias sobre proveedores y eximentes. La aplicación a un servicio mediado por IA exige examinar el texto vigente, la calidad de consumidor, el defecto alegado y la conexión causal; no basta con importar la solución europea.

En cualquier jurisdicción, la calificación como producto o servicio puede ser decisiva. Un modelo descargable, una aplicación accesible en la nube, una herramienta integrada en un dispositivo médico y un servicio profesional que usa IA no necesariamente reciben el mismo tratamiento legal. La arquitectura comercial puede combinar elementos: software

suministrado como producto, mantenimiento continuo como servicio y decisiones tomadas por un profesional. El régimen aplicable puede ser acumulativo, pero deben evitarse duplicaciones indemnizatorias por el mismo daño.

7.1.4. Contrato y responsabilidad frente a terceros

Los contratos distribuyen obligaciones entre proveedor, integrador, operador y cliente institucional. Pueden asignar deberes de prueba, monitoreo, actualización, seguridad, notificación de incidentes y cooperación en litigios. También suelen incluir límites de responsabilidad, indemnidades y cláusulas sobre propiedad o uso de datos. Estas disposiciones pueden ordenar la relación interna entre las partes, pero no necesariamente son oponibles a terceros ni pueden eliminar derechos imperativos de consumidores o personas afectadas.

Una cláusula que atribuye todos los riesgos al cliente puede resultar insuficiente si el proveedor conserva control sobre el modelo, decide las actualizaciones o retiene información esencial para investigar una falla. De manera similar, una entidad pública no debería trasladar íntegramente al proveedor su deber de fundamentar decisiones administrativas o garantizar un recurso efectivo. La asignación contractual debe corresponder a capacidades reales de control e información, y no usarse para crear una zona de impunidad.

La relación entre responsabilidad contractual y extracontractual también depende de la jurisdicción. El usuario que contrató el servicio puede reclamar incumplimiento; una persona ajena al contrato podría invocar daños extracontractuales o normas de consumo si reúne sus requisitos. Los acuerdos entre empresas pueden prever indemnización cruzada, pero la reparación de la víctima se rige por la ley aplicable y por las reglas de legitimación y causalidad.

La responsabilidad civil, por último, no se limita a daños patrimoniales.

Dependiendo del ordenamiento, pueden reconocerse daños personales, morales, reputacionales o pérdida de oportunidades. La cuantificación requiere relacionar la afectación con el sistema y distinguir el daño indemnizable de las consecuencias hipotéticas. En el caso de daños colectivos o repetidos, las acciones colectivas y mecanismos de representación pueden ser relevantes cuando la ley los permite.

7.2. Atribución de responsabilidad entre desarrolladores, proveedores, desplegadores y usuarios

7.2.1. La cadena de suministro como unidad de análisis

Un sistema de IA no suele ser obra de un solo actor. Un desarrollador puede crear el modelo; un proveedor lo comercializa; un integrador lo adapta; una organización lo despliega; un profesional lo utiliza; y un subcontratista procesa los datos o mantiene la infraestructura. La distribución de responsabilidades debe considerar la cadena completa, pues el daño puede deberse a una deficiencia de diseño, a una modificación posterior, a un contexto de uso inadecuado o a la combinación de varias decisiones.

Conviene distinguir dos preguntas. La primera es quién tiene deberes jurídicos frente a la persona afectada. La segunda es cómo se reparte finalmente el costo entre los participantes, por contrato o por acción de repetición. La persona damnificada no debería verse obligada a reconstruir desde el inicio todas las relaciones comerciales internas si la ley permite dirigir el reclamo contra un proveedor, empleador, institución o fabricante responsable. La asignación interna puede resolverse posteriormente entre los actores.

El desarrollador suele controlar decisiones sobre arquitectura, entrenamiento, pruebas iniciales y documentación. Su responsabilidad puede vincularse con fallas previsibles de diseño, datos inadecuados, límites no comunicados o ausencia de salvaguardias razonables. Sin embargo, no toda salida perjudicial demuestra un defecto de

desarrollo: la aplicación puede haber cambiado de propósito, población o condiciones operativas.

El proveedor que comercializa el sistema puede asumir deberes de información, soporte, actualización, seguridad y respuesta a incidentes. En servicios de IA como servicio, puede mantener control continuo sobre el modelo y sus cambios. La responsabilidad se valora según el rol efectivamente desempeñado y la normativa aplicable; el nombre que se dé en el contrato no siempre determina la calificación jurídica.

7.2.2. Integrador, entidad desplegadora y usuario profesional

El integrador combina el sistema con bases de datos, software, interfaces o procesos existentes. Puede modificar el modelo, ajustar umbrales, agregar funciones o conectar sus resultados con decisiones automáticas posteriores. Si la integración altera sustancialmente el comportamiento o el propósito, el integrador puede asumir deberes propios de evaluación y documentación, según la legislación aplicable. Si únicamente instala el sistema bajo instrucciones precisas, su posición será distinta.

La entidad desplegadora decide para qué se utiliza la herramienta y quiénes serán afectados. Esa entidad suele controlar variables esenciales: selección del proveedor, definición del propósito, calidad de los datos locales, supervisión humana, configuración del flujo de trabajo y respuesta ante errores. En contextos laborales, financieros, educativos, sanitarios o administrativos, desplegar una herramienta sin validar su adecuación al uso concreto puede constituir una conducta jurídicamente relevante, incluso cuando el software funcione como fue diseñado.

El usuario profesional —por ejemplo, un médico, abogado, analista o funcionario— puede tener un deber de verificación independiente.

La intensidad de ese deber depende de la competencia profesional, la relevancia de la decisión, la disponibilidad de información y las normas sectoriales. Confiar ciegamente en una salida probabilística no equivale necesariamente a diligencia profesional. A la vez, la organización no puede trasladar todo el riesgo al operador individual si el sistema está diseñado para inducir confianza, si el tiempo de revisión es insuficiente o si las metas institucionales premian la aceptación automática.

En la práctica, el diseño organizacional puede ser tan importante como el modelo. Si el procedimiento hace que las recomendaciones algorítmicas se conviertan en decisiones definitivas, aunque exista formalmente un botón de "revisión humana", la responsabilidad puede alcanzar a quienes establecieron el flujo. La supervisión debe ser efectiva y documentada: quién revisó, qué información tuvo, qué posibilidad existía de apartarse del resultado y cómo se gestionó la incertidumbre.

7.2.3. Control, conocimiento y beneficio

Tres criterios ayudan a ordenar la atribución: control, conocimiento y beneficio. El control se refiere a la capacidad de diseñar, modificar, detener o supervisar el sistema. El conocimiento comprende información sobre limitaciones, riesgos, incidentes y condiciones de uso. El beneficio considera quién obtiene ventajas económicas u organizacionales del despliegue. Ninguno de estos factores es por sí solo una regla universal de responsabilidad, pero su combinación puede orientar la identificación de deberes y la distribución de costos.

Un proveedor puede controlar actualizaciones y tener datos de desempeño que el cliente no conoce. La entidad que despliega puede decidir usar el sistema para filtrar solicitudes de crédito y observar sus efectos sobre su clientela. El profesional puede intervenir en el resultado final, pero carecer de herramientas para auditar el modelo. La atribución debe evitar dos extremos: responsabilizar siempre al proveedor sin importar el uso, o imputar todo al usuario final aunque

no controle ni comprenda el sistema.

La idea de una "brecha de responsabilidad" es útil para analizar casos en los que cada actor señala al siguiente y la persona afectada no puede obtener reparación. No significa que deba reconocerse personalidad jurídica a la IA. Dotar al sistema de personalidad no necesariamente crea un patrimonio disponible, una capacidad de reparar o un mecanismo real de prevención. Puede incluso desplazar la atención de actores humanos y corporativos que sí obtienen beneficios y pueden adoptar medidas de cuidado.

La doctrina debe diferenciar responsabilidad moral, responsabilidad organizacional y responsabilidad jurídica. Una empresa puede tener responsabilidad de gobernanza por fallas sistémicas; una persona puede responder por una acción u omisión concreta; una autoridad puede ser responsable por la decisión pública que adopta. El derecho debe determinar qué consecuencias corresponden a cada nivel sin diluir la imputación individual cuando exista conducta culpable ni fragmentar tanto la cadena que nadie responda.

7.2.4. Modificaciones, datos y usos fuera de propósito

La actualización de un sistema plantea cuestiones de atribución temporal. Una herramienta puede ser razonablemente segura al momento de su lanzamiento y volverse defectuosa tras una actualización; o puede mejorar en promedio, pero degradarse para un grupo específico. El análisis debe determinar quién decidió el cambio, quién tuvo oportunidad de evaluarlo, qué avisos se emitieron y si el despliegue incorporó una nueva finalidad que requería una evaluación distinta.

El uso fuera de propósito también exige matices. Si un cliente utiliza un modelo de generación de texto para una función no prevista, la responsabilidad puede recaer en quien modificó el sistema o lo aplicó

de manera imprudente. Pero el proveedor puede tener obligaciones si el uso indebido era previsible y pudo mitigarse mediante advertencias, controles de acceso o diseño seguro. La previsibilidad y la posibilidad de prevención deben examinarse con evidencia, no mediante presunciones abstractas.

La calidad de los datos puede originar responsabilidades diferenciadas. Quien recopila o selecciona datos podría incumplir normas de protección de datos o deberes contractuales; quien entrena el modelo podría no detectar sesgos previsibles; quien despliega podría alimentar al sistema con información desactualizada o inexacta. Como se explicó en los capítulos anteriores, el sesgo algorítmico y la opacidad son factores relevantes, pero no reemplazan la prueba del daño ni determinan automáticamente quién debe indemnizar.

Los mecanismos de notificación y cooperación son esenciales para reconstruir la cadena causal. Contratos y marcos regulatorios pueden exigir conservar registros, comunicar incidentes, permitir auditorías y mantener documentación de versiones. Sin estos deberes, el sistema puede cambiar mientras avanza el litigio, desaparecer el registro técnico o perderse la posibilidad de determinar si el daño provino de un defecto, una mala integración o una decisión humana.

Tabla 7.1

Actores de la cadena de IA y elementos para analizar su responsabilidad

Actor	Función o ámbito de control	Elementos que conviene revisar	Pregunta para el análisis jurídico
Desarrollador	Diseña, entrena o prueba el sistema	Datos y métodos utilizados, pruebas, limitaciones conocidas y advertencias	¿Identificó y comunicó riesgos previsibles durante el diseño y desarrollo?

Actor	Función o ámbito de control	Elementos que conviene revisar	Pregunta para el análisis jurídico
Proveedor	Comercializa o pone el sistema a disposición de terceros	Documentación, instrucciones, condiciones de uso, soporte y actualizaciones	¿Informó sobre el uso previsto, las limitaciones y las medidas de seguridad necesarias?
Integrador	Incorpora el sistema a una aplicación o servicio y puede modificar su configuración	Cambios realizados, interfaces, datos conectados y pruebas de integración	¿Las modificaciones o integraciones alteraron el comportamiento o aumentaron el riesgo?
Entidad desplegadora	Decide la finalidad y el contexto en que se utiliza el sistema	Evaluación previa, selección de usuarios, protocolos, supervisión y registros	¿Utilizó el sistema de forma adecuada al contexto y adoptó controles proporcionales?
Usuario profesional	Interpreta o aplica los resultados dentro de su actividad	Formación, instrucciones recibidas, intervención y decisión final	¿Comprendió los límites del resultado y podía revisarlo o apartarse de él?
Tercero que aporta datos o servicios	Proporciona datos, infraestructura u otros componentes	Procedencia, calidad, condiciones contractuales y trazabilidad	¿Su aporte influyó en el daño y qué capacidad tenía para prevenirlo o corregirlo?

Nota. *La tabla identifica aspectos que pueden orientar el análisis, pero no determina automáticamente quién es responsable. La atribución depende de los hechos, la capacidad de control, la previsibilidad del daño y las normas aplicables. Elaboración propia con base en los temas tratados en el capítulo.*

7.3. Responsabilidad penal, administrativa y sancionatoria

7.3.1. Responsabilidad penal y principio de culpabilidad

La responsabilidad penal exige más que demostrar que una IA produjo un resultado dañino. Deben acreditarse los elementos del tipo penal, la conducta atribuible a una persona o entidad reconocida por el ordenamiento, el elemento subjetivo requerido y las condiciones de imputación. Los sistemas de IA no poseen, en el derecho vigente de la mayoría de jurisdicciones, personalidad penal general ni capacidad propia de culpabilidad. Por ello, no es correcto afirmar que "la IA cometió un delito" en sentido jurídico sin identificar la conducta humana o corporativa penalmente relevante.

En delitos dolosos, la cuestión puede ser si una persona utilizó deliberadamente el sistema para ejecutar una conducta ilícita, conocía el riesgo y quiso producir el resultado o aceptó su posible ocurrencia, según la definición local del dolo. En delitos imprudentes, se examina si el sujeto infringió un deber objetivo de cuidado y si el resultado era previsible y evitable conforme al estándar aplicable. La IA puede ser instrumento, medio de ejecución o elemento de contexto, pero no sustituye la prueba de autoría y culpabilidad.

La utilización de un sistema puede dificultar la previsibilidad. Si una persona configura una herramienta para actuar automáticamente, debe analizarse qué control conservó, qué riesgos conocía y si era razonable confiar en el comportamiento del sistema. La imprevisibilidad técnica no borra automáticamente la responsabilidad, pero tampoco es justo atribuirla con base exclusiva en el resultado. La investigación penal debe identificar la evidencia disponible antes del hecho, las advertencias recibidas, las posibilidades de intervención y las decisiones posteriores.

La responsabilidad penal de las personas jurídicas varía ampliamente. Algunos ordenamientos la reconocen para determinados delitos y bajo requisitos de atribución corporativa; otros mantienen modelos

centrados principalmente en personas naturales. Allí donde existe responsabilidad penal corporativa, no se imputa por el mero uso de IA: deben acreditarse los criterios legales, como actuación en beneficio, defecto de organización o intervención de órganos o representantes, de acuerdo con la ley nacional. En Ecuador y otros países latinoamericanos, cualquier análisis concreto debe verificar el tipo penal, la regulación de personas jurídicas y la jurisprudencia aplicable.

7.3.2. Conducta de terceros y usos criminales

Una persona puede utilizar IA para facilitar fraude, suplantación, extorsión, falsificación, acoso o manipulación de información. En tales casos, la responsabilidad penal se centra en la conducta del autor y en los delitos tipificados por la legislación vigente. La novedad tecnológica no crea por sí sola un delito; el principio de legalidad exige que la conducta esté previamente prevista en la ley. Si una tecnología amplifica la capacidad de causar daño, ello puede incidir en la prueba del dolo, la ejecución o la gravedad, siempre que el ordenamiento lo autorice.

El proveedor no responde penalmente por todo uso ilícito de su herramienta. Para imputarle responsabilidad se requiere demostrar los elementos jurídicos aplicables, incluida la participación, conocimiento, propósito o incumplimiento penalmente relevante. La mera previsibilidad de que una tecnología pueda ser mal utilizada no equivale automáticamente a complicidad. Las obligaciones administrativas de seguridad o diligencia pueden ser relevantes para otros ámbitos, pero no sustituyen la exigencia penal de tipicidad y culpabilidad.

Los problemas de evidencia digital son importantes. Registros, instrucciones, versiones del modelo, cuentas de acceso y metadatos pueden desaparecer o estar en poder de proveedores extranjeros. La investigación debe preservar evidencia con garantías de autenticidad, cadena de custodia y respeto a derechos fundamentales. Si se utilizan herramientas de IA para analizar datos o priorizar sospechosos, la

propia investigación debe poder explicar su rol y someterse a control, particularmente cuando el resultado incide en una medida coercitiva.

7.3.3. Responsabilidad administrativa y sanciones regulatorias

La responsabilidad administrativa puede activarse cuando una entidad incumple deberes regulatorios, aun cuando no se pruebe un daño individual indemnizable. Autoridades de protección de datos, consumo, competencia, trabajo, salud, finanzas o telecomunicaciones pueden sancionar conductas que estén dentro de sus competencias. Entre ellas podrían figurar el tratamiento ilícito de datos, la publicidad engañosa, el incumplimiento de deberes de información, la discriminación prohibida o el uso de un sistema fuera de los límites regulatorios.

Las sanciones administrativas y la reparación civil cumplen funciones distintas. La multa busca disuadir y expresar reproche institucional; la indemnización busca reparar el daño sufrido. Una sanción administrativa no garantiza que la persona reciba compensación, y un acuerdo indemnizatorio no elimina necesariamente el interés público en investigar una infracción. Los procedimientos deben coordinarse para evitar decisiones incompatibles y proteger derechos de defensa, legalidad y proporcionalidad.

El Reglamento de IA de la Unión Europea (Reglamento [UE] 2024/1689) establece obligaciones regulatorias y sanciones administrativas para ciertas infracciones, con aplicación escalonada. Estas multas no son un régimen general de indemnización civil: la persona que busca compensación debe acudir a las reglas de responsabilidad aplicables. La distinción es fundamental, pues cumplir con obligaciones regulatorias no excluye necesariamente toda responsabilidad civil, y una indemnización no sustituye la supervisión del mercado.

En la Unión Europea, la propuesta de Directiva sobre responsabilidad

extracontractual adaptada a la IA —conocida como *AI Liability Directive*— fue retirada del procedimiento legislativo. La propuesta había contemplado mecanismos sobre exhibición de pruebas y presunciones de causalidad, pero no debe citarse como norma vigente. La Directiva (UE) 2024/2853 sobre productos defectuosos sí fue aprobada y aborda software, aunque tiene un alcance diferente y una fecha de transposición propia. Este desarrollo demuestra que regulación de sistemas y responsabilidad por daños son campos relacionados, pero no intercambiables. **7.3.4 Sanción, prevención y garantías**

Una sanción adecuada debe tener base legal, ser proporcional a la gravedad, considerar la conducta y permitir revisión independiente. En materia de IA, la autoridad puede necesitar valorar el número de personas afectadas, la duración del uso, la sensibilidad del contexto, el nivel de control del infractor, la cooperación posterior y la repetición de incidentes. Sin embargo, esos criterios deben encontrarse en la norma aplicable o derivarse de ella; no pueden inventarse por analogía para agravar una sanción.

Las medidas correctivas pueden ser tan importantes como la multa. La autoridad podría ordenar suspender un tratamiento, corregir datos, revisar decisiones, informar a las personas afectadas o retirar un sistema cuando la ley habilita esa actuación. Estas medidas deben dirigirse a detener el riesgo y restablecer garantías, no convertirse en castigos encubiertos sin debido proceso.

Las organizaciones también pueden adoptar controles internos de cumplimiento: mapa de sistemas, asignación de responsables, evaluación de impacto, revisión de proveedores, canal de incidentes y capacitación. La existencia de un programa de cumplimiento no exime automáticamente de responsabilidad, pero puede mostrar medidas preventivas y facilitar detección temprana. Para que sea significativo, el programa debe tener independencia, recursos y capacidad de detener el despliegue.

La coordinación entre sanción administrativa, responsabilidad civil y proceso penal debe preservar la presunción de inocencia y evitar doble sanción incompatible con las reglas de cada país. Las autoridades deberían compartir información con salvaguardias, facilitar el acceso de las víctimas a evidencia pertinente y evitar que los procedimientos se fragmenten hasta hacer imposible una respuesta integral.

7.4. Seguros, carga de la prueba y mecanismos de reparación

7.4.1. Causalidad y opacidad técnica

Una víctima que reclama indemnización suele tener que probar que el daño se relaciona jurídicamente con la conducta o producto demandado. En sistemas de IA, la causalidad puede ser difícil de establecer por tres motivos. Primero, la salida depende de múltiples variables y procesos intermedios. Segundo, el sistema puede cambiar con actualizaciones o con el contexto de uso. Tercero, la documentación y los registros suelen estar bajo control del proveedor o de la organización que lo utiliza.

La opacidad no es una característica uniforme. Algunos sistemas son técnicamente interpretables; otros son complejos, protegidos por secretos comerciales o inaccesibles para la persona afectada. Incluso cuando el modelo puede describirse, no siempre es posible reconstruir por qué produjo una salida específica en un momento determinado. El estudio del Parlamento Europeo sobre responsabilidad civil y sistemas de IA subraya que esa complejidad puede dificultar identificar los pasos que llevaron de una entrada a una salida y, con ello, la prueba del defecto o de la causalidad.

La dificultad probatoria puede generar asimetría. Una persona recibe una decisión adversa y desconoce que intervino un algoritmo; la institución conserva el expediente, el registro de puntuación y el contrato; el proveedor posee los datos de rendimiento y documentación técnica. Exigir a la persona que identifique desde el inicio el mecanismo interno

puede hacer ilusorio el derecho a reclamar. Por ello, la transparencia procesal, la preservación de evidencia y los deberes de cooperación son esenciales.

La prueba debe diferenciar correlación, contribución y causa jurídicamente suficiente. Que una herramienta haya intervenido no demuestra que causó el daño; pero tampoco debe exigirse reconstruir cada parámetro del modelo si existen indicios sólidos de que la decisión automatizada fue determinante. El estándar causal aplicable depende de la jurisdicción y del tipo de responsabilidad. Los jueces pueden considerar evidencia estadística y pericial, sin convertirla en sustituto automático del análisis individual.

7.4.2. Acceso a la evidencia y carga probatoria

Los ordenamientos cuentan con herramientas procesales que pueden ayudar: exhibición de documentos, peritajes, presunciones, inversión de carga en ciertos casos, medidas cautelares y reglas de protección al consumidor. Su disponibilidad y alcance varían. La opacidad de un modelo no autoriza a un juez a crear una presunción sin base legal, pero puede justificar interpretar de manera efectiva las reglas de acceso a prueba y valorar quién tenía control de la información.

Las empresas y entidades que utilizan sistemas deberían conservar registros que permitan responder preguntas básicas: qué versión del sistema estaba operativa, qué entradas relevantes se procesaron, qué salida se generó, quién la revisó, qué cambios se hicieron y qué advertencias estaban disponibles. La documentación debe conservarse de forma segura y proporcional, con límites de retención que respeten la protección de datos.

La Unión Europea intentó abordar parte de esta asimetría mediante la propuesta retirada de AI Liability Directive. Su retirada significa que las reglas procesales propuestas no constituyen derecho vigente

de la Unión. En cambio, la Directiva 2024/2853 sobre productos defectuosos sí contempla herramientas de divulgación y presunciones en ciertos supuestos, cuyo alcance debe leerse dentro de ese régimen de producto y conforme a su transposición nacional. No es correcto generalizar esas disposiciones como una regla probatoria universal para toda reclamación relacionada con IA.

El peritaje puede explicar el funcionamiento, las limitaciones y el desempeño del sistema, pero la experticia técnica no decide por sí sola la imputación jurídica. El juez debe examinar si la conducta fue diligente, si el producto ofrecía seguridad adecuada, si la decisión respetó normas aplicables y si el nexo causal satisface el estándar legal. También es necesario proteger la contradicción: la parte afectada debe poder cuestionar el método, los datos y las inferencias del perito.

7.4.3. Seguros y distribución económica del riesgo

El seguro puede facilitar la reparación y distribuir riesgos entre múltiples incidentes. Las pólizas de responsabilidad profesional, ciberseguridad, productos y responsabilidad general pueden cubrir ciertos daños vinculados a IA, según sus términos. No existe una cobertura uniforme: algunas pólizas excluyen pérdidas derivadas de decisiones algorítmicas, infracciones de propiedad intelectual, discriminación o fallas de seguridad. La cobertura debe revisarse con atención a la actividad asegurada, el tipo de daño, el ámbito territorial, las exclusiones y los límites.

La existencia de una póliza no sustituye la prevención. Si los aseguradores pueden evaluar controles, historial de incidentes y calidad de gobernanza, pueden incentivar prácticas más seguras mediante condiciones y primas. Pero las aseguradoras pueden tener información insuficiente para valorar modelos complejos o riesgos correlacionados. Una falla sistémica puede afectar simultáneamente a miles de personas, lo que desafía esquemas diseñados para eventos individuales y relativamente independientes.

Para pequeñas empresas y entidades públicas, el costo de asegurar ciertos sistemas puede limitar su adopción. Esa tensión no debe resolverse eliminando derechos de las víctimas. Puede requerir fondos sectoriales, seguros obligatorios para actividades de alto riesgo cuando la ley lo disponga, garantías financieras, límites de cobertura o mecanismos colectivos. Cualquier modelo de este tipo necesita evaluación económica y normativa para evitar trasladar costos a quienes tienen menor capacidad de pago.

La Directiva europea de productos defectuosos no crea un seguro obligatorio general para la IA. La propuesta de responsabilidad civil de IA tampoco llegó a convertirse en legislación. Por ello, las decisiones sobre seguros y garantías siguen dependiendo de normas sectoriales, contratos y regímenes nacionales. En América Latina, la discusión debería partir de los mercados aseguradores existentes, la capacidad de supervisión y la necesidad de asegurar reparación efectiva en actividades de alto impacto.

7.4.4. Reparación integral y mecanismos preventivos

La reparación debe guardar relación con la naturaleza del daño y las reglas del ordenamiento. Puede comprender indemnización económica, restitución, corrección de datos, revisión de una decisión, reincorporación a un servicio, rectificación pública u otras medidas, según el caso. Cuando una decisión automatizada genera exclusión reiterada, pagar una suma individual puede no resolver la causa estructural. La reparación puede requerir cambiar el sistema, revisar decisiones anteriores o establecer garantías para evitar repetición.

Los mecanismos extrajudiciales pueden ofrecer respuestas más rápidas: procedimientos ante autoridades de consumo o protección de datos, defensorías, mediación, ombudsman institucional y canales internos de reclamación. Estos mecanismos deben ser accesibles, no exigir conocimiento técnico avanzado y no impedir el acceso posterior a tribunales. También requieren plazos claros, independencia,

motivación de las decisiones y posibilidad de obtener evidencia.

En daños masivos, las acciones colectivas, demandas representativas y procedimientos administrativos pueden ser relevantes para agrupar reclamos similares, siempre que la ley nacional los contemple. La naturaleza repetida de ciertas decisiones algorítmicas puede hacer que cada persona afectada sufra un perjuicio individual relativamente pequeño, mientras que el impacto conjunto sea considerable. La agregación puede reducir costos procesales y permitir examinar patrones que no se aprecian en casos aislados.

La prevención es parte de una política de responsabilidad eficaz. Las evaluaciones de impacto, auditorías, registros, pruebas previas y monitoreo pueden reducir daños antes de que se produzcan. Pero no debe confundirse la prevención con inmunidad. Cumplir procedimientos no excluye toda responsabilidad si persiste un defecto o una conducta negligente; del mismo modo, un resultado adverso no demuestra automáticamente que el sistema fuera ilegal.

En conclusión, la responsabilidad por daños causados por sistemas de IA debe asignarse a actores jurídicos capaces de prevenir, controlar, informar y reparar. Los regímenes tradicionales de culpa, riesgo, producto, contrato y sanción siguen ofreciendo herramientas, aunque la opacidad y la distribución del control hacen necesarias reglas procesales y prácticas de documentación más sólidas. La IA no elimina la responsabilidad humana u organizacional: obliga a precisar quién tenía qué deber, qué información poseía, qué podía hacer para evitar el daño y qué reparación requiere la persona afectada.

Figura 7.1.

Análisis causal y reparación de un daño relacionado con IA



Nota: *Elaboración propia. Hay que señalar que la causalidad y la reparación dependen de las normas aplicables y de las circunstancias del caso.*



Capítulo 8.

Inteligencia artificial en la administración pública y la justicia



Introducción

El uso de inteligencia artificial (IA) por el Estado plantea preguntas jurídicas particulares: ¿puede una administración automatizar una decisión que afecta derechos?, ¿qué garantías debe ofrecer un tribunal que utiliza sistemas de apoyo?, ¿qué límites rigen la predicción de riesgos en seguridad pública? En estos ámbitos, la eficiencia no basta para justificar una herramienta. También deben considerarse la legalidad, la motivación, la igualdad, el debido proceso y la posibilidad de impugnar decisiones.

Este capítulo estudia cuatro espacios relacionados, pero distintos: las decisiones administrativas automatizadas; la IA en la actividad judicial; su empleo en seguridad y control penal; y los mecanismos de transparencia y control judicial. Se distingue entre automatizar tareas auxiliares y delegar decisiones que corresponden a autoridades humanas, así como entre los marcos jurídicos vinculantes y las orientaciones éticas o administrativas.

8.1. Decisiones administrativas automatizadas y debido proceso algorítmico

8.1.1. La decisión pública y la intervención algorítmica

La administración pública utiliza sistemas automatizados para clasificar solicitudes, detectar inconsistencias, priorizar inspecciones, asignar recursos, estimar riesgos o recomendar decisiones. Estas herramientas pueden agilizar tareas repetitivas y ayudar a procesar grandes volúmenes de información. Sin embargo, cuando su resultado influye en la concesión de un beneficio, la imposición de una sanción, el acceso a un servicio o la evaluación de una persona, la decisión conserva carácter público, aunque haya sido preparada por un algoritmo o por un proveedor privado.

Es necesario distinguir entre tres niveles de intervención. En el primero, la IA cumple una función auxiliar: organiza documentos, extrae datos o sugiere información para el funcionario. En el segundo, ofrece una recomendación o puntuación que influye de manera importante en la decisión humana. En el tercero, determina automáticamente el resultado o activas consecuencias sin una revisión humana sustantiva. El nivel de automatización no decide por sí solo la legalidad, pero sí modifica las garantías necesarias y el grado de escrutinio requerido.

La autoridad no puede eludir sus obligaciones alegando que el resultado fue producido por un sistema externo. Si una entidad adopta una herramienta para ejercer una competencia pública, conserva la responsabilidad de actuar dentro de la ley, motivar sus decisiones y respetar los derechos de las personas. Un contrato de adquisición no transforma una decisión administrativa en una decisión privada ni desplaza la competencia legal del órgano.

Por ello, la adopción de IA debe partir de una pregunta institucional: ¿qué problema público se pretende resolver y por qué un sistema automatizado resulta apropiado? La disponibilidad de una herramienta

no constituye, por sí misma, una razón suficiente para emplearla. Deben considerarse alternativas menos intrusivas, la calidad de los datos, el impacto sobre distintos grupos, las posibilidades de error y la capacidad de la administración para supervisar el resultado.

8.1.2. Legalidad, motivación y revisión humana

El principio de legalidad requiere que la administración actúe dentro de sus competencias y con base en normas aplicables. Cuando una decisión automatizada afecta derechos, la entidad debe poder explicar quién la adoptó, qué competencia ejerció y cómo el sistema participó en el procedimiento. La automatización no puede crear facultades que la ley no concede ni reducir garantías que el procedimiento administrativo reconoce.

La motivación debe permitir comprender las razones relevantes del acto. No siempre será necesario revelar el código fuente ni exponer todos los parámetros técnicos. Sí debe ser posible conocer los hechos considerados, los criterios decisivos, el papel del sistema y el modo en que la autoridad evaluó su resultado. Una explicación genérica —por ejemplo, "el sistema determinó que la solicitud no cumple"— no permite a la persona identificar un error ni formular una impugnación efectiva.

La intervención humana debe ser real cuando el marco jurídico o la gravedad de la decisión así lo exijan. La persona que revisa necesita competencia para modificar el resultado, acceso a información suficiente, tiempo y capacitación. Si el funcionario solo confirma automáticamente una puntuación sin analizar el expediente, existe supervisión nominal, pero no control sustantivo. La organización también debe evitar que sus metas de productividad presionen al personal para aceptar las recomendaciones del sistema sin examinarlas.

El debido proceso algorítmico puede incluir información sobre el uso

de la herramienta, acceso a los elementos relevantes del expediente, oportunidad para corregir datos inexactos, revisión por una autoridad competente y decisión motivada. La intensidad de las garantías dependerá de la naturaleza de la medida. Una recomendación para ordenar turnos administrativos no tiene el mismo peso que una herramienta que contribuye a suspender una prestación, denegar una licencia o determinar la elegibilidad de una persona para recibir asistencia.

La revisión debe ser accesible y oportuna. Si la impugnación llega después de que la persona perdió una prestación esencial, la posibilidad de recurrir puede resultar insuficiente. En los casos en que exista riesgo de daño grave o difícilmente reparable, el ordenamiento puede requerir medidas provisionales o una revisión prioritaria. La existencia de un recurso formal no garantiza su eficacia si la persona desconoce que hubo automatización o no puede obtener la información necesaria para cuestionarla.

8.1.3. Igualdad, datos y servicios esenciales

Los sistemas administrativos pueden reproducir desigualdades presentes en los datos o en la organización que los utiliza. Por ejemplo, un modelo que prioriza inspecciones a partir de registros históricos podría concentrarlas en barrios ya sometidos a una vigilancia más intensa. Una herramienta para detectar fraude podría penalizar a solicitantes cuyas trayectorias laborales o familiares difieren de las mayoritarias en los datos de entrenamiento. La administración debe evaluar si sus criterios son pertinentes, si las variables actúan como sustitutos de características protegidas y si los errores se distribuyen de manera desigual.

La calidad de los datos es una condición jurídica y operativa. Un dato desactualizado, una identidad mal asociada o una categoría incompleta puede producir un resultado adverso. La persona afectada debe poder solicitar corrección y la entidad debe mantener procedimientos para

comprobar si la base utilizada es adecuada para su finalidad. Como se desarrolló en los capítulos 4 y 5, la protección de datos y el análisis de sesgos son componentes de esta evaluación, pero no reemplazan el examen de competencia, motivación y recurso.

El Reglamento de IA de la Unión Europea incluye como sistemas de alto riesgo ciertos usos de IA por autoridades para evaluar la elegibilidad de personas para servicios y prestaciones públicas esenciales, así como determinados usos en seguridad y justicia. Esa clasificación establece obligaciones regulatorias para los sistemas comprendidos en su ámbito; no convierte automáticamente toda herramienta pública en sistema de alto riesgo ni sustituye las garantías del derecho administrativo. La ley aplicable, el uso concreto y las excepciones previstas deben revisarse en cada caso.

La administración también debe considerar accesibilidad y no discriminación indirecta. Si el trámite automatizado solo puede completarse mediante una aplicación, una persona con discapacidad, sin conectividad adecuada o con dificultades lingüísticas podría quedar excluida. La existencia de un canal humano alternativo, asistencia adecuada y medios para aportar evidencia puede ser necesaria para que la digitalización no se convierta en una barrera de acceso a derechos.

8.1.4. Contratación, documentación y responsabilidad institucional

Las administraciones suelen adquirir herramientas de proveedores que conservan control sobre el modelo, su mantenimiento o sus actualizaciones. Los contratos deben permitir que la entidad compradora conozca las limitaciones relevantes, supervise el desempeño, audite usos autorizados, conserve registros y comunique incidentes. Las cláusulas de confidencialidad comercial no deberían impedir que la autoridad explique una decisión ni que los órganos de control investiguen un posible incumplimiento.

También es importante documentar el ciclo de vida del sistema. Antes de contratarlo, la entidad debería definir la finalidad, identificar a las personas afectadas y evaluar alternativas. Durante la implementación, debe probar la herramienta en condiciones que reflejen el contexto real. Después del despliegue, necesita monitorear fallos, revisar cambios y establecer criterios para suspender o retirar el sistema. La documentación facilita tanto la supervisión preventiva como la respuesta a reclamaciones.

Brasil ofrece un ejemplo de instrumentos administrativos para el sector público. La Secretaría de Gobierno Digital publicó guías sobre IA generativa y un Guía Unificado de IA para orientar proyectos desde su planificación hasta su desactivación. Son herramientas de apoyo a la gestión federal, no una ley general que cree por sí sola derechos individuales frente a toda decisión automatizada. Su interés reside en organizar responsabilidades y buenas prácticas a lo largo del ciclo de vida.

La documentación también debe sostener una distribución clara de responsabilidades dentro del organismo. Debe identificarse qué área autoriza el uso, qué funcionario responde por la decisión, quién monitorea el sistema y qué unidad atiende reclamos. Si toda responsabilidad queda atribuida de forma abstracta a "la tecnología", la estructura institucional facilita que los errores persistan y que la persona afectada no encuentre una vía concreta de reparación.

8.2. IA en la administración de justicia: justicia predictiva y apoyo judicial

8.2.1. Usos judiciales y tareas auxiliares

En los sistemas judiciales, la IA puede apoyar la gestión de expedientes, la búsqueda de precedentes, la anonimización de documentos, la transcripción, la traducción, la clasificación de escritos o la identificación de asuntos relacionados. También se proponen

herramientas de análisis predictivo que estiman la duración de procesos o identifican tendencias en decisiones previas. Estas aplicaciones no son equivalentes entre sí: algunas automatizan tareas administrativas; otras pueden influir directamente en la interpretación jurídica o en el sentido de una resolución.

La distinción entre apoyo y decisión es central. Un buscador avanzado puede ayudar a localizar precedentes, pero no determina qué regla debe aplicarse. Un sistema que resume un expediente puede ahorrar tiempo, pero puede omitir hechos relevantes. Una herramienta que propone un resultado a partir de sentencias anteriores puede influir en la deliberación incluso si el juez conserva formalmente la potestad de decidir. Por eso, el efecto real sobre el juicio debe analizarse, no solo la categoría comercial del producto.

El uso de IA puede mejorar la eficiencia y accesibilidad de la justicia, especialmente cuando reduce tareas repetitivas y permite que jueces y funcionarios concentren tiempo en análisis que requieren valoración jurídica. No obstante, la eficiencia no puede medirse solo por el número de expedientes tramitados. También importa la calidad de las decisiones, la igualdad de trato, la oportunidad de defensa, la privacidad de quienes participan y la confianza pública en la independencia judicial.

La Comisión Europea para la Eficiencia de la Justicia (CEPEJ) del Consejo de Europa adoptó en 2018 una Carta Ética sobre el uso de IA en sistemas judiciales. Sus principios incluyen respeto a los derechos fundamentales, no discriminación, calidad y seguridad, transparencia e imparcialidad, y control por parte del usuario. La Carta es un instrumento de orientación ética y técnica, no una norma directamente vinculante para todos los tribunales. Aun así, ofrece un marco útil para diseñar políticas judiciales responsables.

8.2.2. Independencia judicial y justicia predictiva

La independencia judicial exige que las decisiones sean adoptadas por jueces sometidos a la Constitución y la ley, no por sistemas de puntuación que sustituyan la valoración jurisdiccional. Las herramientas predictivas pueden describir patrones pasados, pero no determinan por sí mismas el resultado jurídicamente correcto de un caso. Las sentencias históricas pueden reflejar desigualdades, cambios legislativos o criterios superados. Convertirlas en predicciones normativas podría consolidar prácticas previas sin analizar su legitimidad.

La "justicia predictiva" puede referirse a tareas distintas: estimar la duración probable de un proceso, clasificar litigios, calcular probabilidades de éxito o predecir decisiones judiciales. Las primeras pueden servir a la planificación administrativa; las últimas pueden alterar la estrategia de las partes o ejercer presión sobre quien juzga. Si una herramienta predice que cierta demanda tiene baja probabilidad de éxito, una parte con menos recursos podría desistir, aunque su reclamo tenga fundamento jurídico. El impacto no depende solo de que el algoritmo dicte sentencia, sino también de cómo influye en las decisiones humanas.

La autonomía judicial exige preservar la facultad de apartarse de una recomendación automatizada y motivar la decisión con razones jurídicas. Un juez no debería tratar una predicción como prueba concluyente ni como sustituto de la valoración de hechos, normas y argumentos. Cuando se utiliza un sistema de apoyo, debe poder explicar qué función cumplió y qué límites tiene, sobre todo si la herramienta influyó significativamente en una resolución que afecta libertad, patrimonio, familia u otros derechos.

La jurisprudencia de Wisconsin en *State v. Loomis* (2016) ilustra algunos de estos riesgos. El tribunal permitió que un juez considerara una evaluación COMPAS en la etapa de sentencia, pero estableció cautelas: el puntaje no podía ser el factor determinante para decidir encarcelamiento o gravedad de la pena, y debían reconocerse las

limitaciones del sistema. La decisión no creó una regla universal ni validó el uso de evaluaciones de riesgo en todas las jurisdicciones; ofrece un ejemplo de cómo la transparencia, el alcance de la herramienta y la fundamentación independiente pueden incidir en el control judicial.

8.2.3. Asistencia a jueces, abogados y usuarios

Las herramientas generativas plantean riesgos adicionales cuando producen resúmenes, borradores, referencias legales o argumentos. Pueden inventar citas, combinar normas de distintas jurisdicciones o presentar como vigente una regla derogada. El problema no se limita a la inexactitud técnica: presentar ante un tribunal información falsa puede afectar el contradictorio, consumir recursos judiciales y perjudicar a una parte. Quien utiliza estos sistemas en el ejercicio profesional conserva el deber de verificar cada cita y afirmación relevante.

Los tribunales también deben establecer reglas de uso interno. Puede ser legítimo utilizar IA para tareas administrativas o para generar borradores bajo supervisión, siempre que se protejan datos reservados y que la decisión final se adopte con revisión humana. En cambio, cargar expedientes confidenciales en servicios públicos sin garantías contractuales o de seguridad puede comprometer la privacidad, el secreto profesional o la reserva procesal.

La asistencia tecnológica debe diseñarse para reducir barreras, no para crear una justicia de dos velocidades. Herramientas de traducción, interfaces accesibles y sistemas de consulta pueden facilitar el acceso a información; pero la automatización de formularios o la orientación jurídica no debe dar a entender que constituye asesoría individual si no revisa las circunstancias del caso. Las personas deben conocer los límites del sistema y disponer de un canal para solicitar asistencia humana.

La CEPEJ ha desarrollado herramientas para operacionalizar su Carta

y para registrar información sobre sistemas utilizados en el entorno judicial. En 2025, su órgano asesor publicó trabajo dirigido a ayudar a los poderes judiciales a evaluar aplicaciones y compartir información sobre los casos de uso. Estas iniciativas subrayan la importancia de que los tribunales conozcan qué sistemas utilizan y puedan examinarlos de forma periódica.

8.2.4. Gobernanza judicial: el caso de Brasil

Brasil ha desarrollado un marco específico para el uso de IA en su Poder Judicial. La Resolución n.º 615/2025 del Consejo Nacional de Justicia establece disposiciones sobre desarrollo, gobernanza, auditoría, monitoreo y uso responsable de soluciones de IA. También prevé el registro de soluciones en la plataforma Sinapses y la clasificación de sistemas según niveles de riesgo. En 2026, el texto compilado registró modificaciones posteriores, lo que muestra la necesidad de verificar la versión vigente de la norma antes de aplicarla.

La norma brasileña reconoce que la IA puede ser utilizada para mejorar la eficiencia del servicio judicial, pero subordina su adopción a derechos fundamentales y a mecanismos de auditoría y monitoreo. El Comité Nacional de Inteligencia Artificial del Poder Judicial funciona como instancia de gobernanza y seguimiento. Este modelo muestra cómo un órgano judicial puede establecer reglas internas vinculantes para sus tribunales sin esperar a que exista una ley general de IA aplicable a todos los sectores.

Su evaluación debe atender a la implementación concreta. Un registro es útil si permite identificar las herramientas y sus usos; una clasificación de riesgos aporta valor si determina controles proporcionales; y una auditoría es significativa si existe acceso a la información necesaria y capacidad para ordenar correcciones. La gobernanza judicial debe preservar autonomía jurisdiccional, impedir que la tecnología condicione el contenido de las sentencias y hacer posible que las partes conozcan cuándo una herramienta intervino de manera relevante.

8.3. IA en seguridad, policía predictiva y control penal

8.3.1. Predicción y vigilancia policial

Las fuerzas de seguridad utilizan o exploran sistemas para analizar imágenes, reconocer identidades, priorizar patrullajes, vincular registros, detectar patrones o estimar riesgos. La etiqueta "policía predictiva" suele abarcar tecnologías diferentes: algunas identifican áreas geográficas con concentración histórica de delitos; otras generan puntuaciones individuales. La diferencia importa, pues una herramienta que orienta la distribución de recursos en un territorio plantea problemas distintos de un sistema que clasifica personas como sospechosas o peligrosas.

Los sistemas basados en datos históricos pueden reproducir ciclos de vigilancia. Si una zona fue objeto de más patrullajes y controles, es probable que sus registros contengan más incidentes que los de territorios menos observados. Un modelo que interpreta esos datos como indicador neutral puede recomendar todavía más vigilancia en los mismos lugares. El resultado puede reforzar desigualdades, reducir la confianza comunitaria y aumentar interacciones coercitivas con poblaciones ya sometidas a un control intensivo.

La predicción no sustituye los requisitos legales para registrar, detener, vigilar o investigar a una persona. Una puntuación de riesgo no debería convertirse por sí sola en causa suficiente para una intervención coercitiva. Las autoridades deben precisar qué información adicional se exige, qué funcionario adopta la medida y cómo se evita que un indicador estadístico sea interpretado como prueba de conducta individual.

El reconocimiento biométrico remoto plantea riesgos especialmente intensos. Puede permitir identificar o rastrear personas sin que estas conozcan cuándo se recopilaron sus datos ni cómo impugnar un error. Los falsos positivos pueden tener consecuencias graves si conducen a

detenciones, interrogatorios o restricciones. El uso policial requiere base legal clara, límites de finalidad, evaluación de precisión, mecanismos de auditoría y vías para rectificar identificaciones erróneas.

8.3.2. Marco regulatorio y prohibiciones

El Reglamento de IA de la Unión Europea trata ciertos usos de IA en seguridad y aplicación de la ley como de alto riesgo e incluye prohibiciones específicas, excepciones y salvaguardias para determinados usos biométricos. También incorpora sistemas destinados a valorar riesgos de infracción o reincidencia en el ámbito de la aplicación de la ley, bajo condiciones definidas. La aplicación debe interpretarse con precisión: no toda herramienta policial está prohibida ni toda herramienta queda autorizada por clasificarse como alto riesgo. La legislación sectorial y las garantías nacionales continúan siendo relevantes.

La clasificación como alto riesgo activa obligaciones regulatorias para proveedores y desplegadores, pero no elimina el control de legalidad de cada medida policial. Si el ordenamiento prohíbe una práctica o no habilita una injerencia, la conformidad técnica del sistema no la vuelve lícita. A la inversa, si el uso está permitido, las autoridades deben demostrar que se implementa con garantías y controles adecuados al derecho fundamental afectado.

La base de datos de entrenamiento, el objetivo operativo y la población de referencia deben examinarse antes de contratar una herramienta. Los resultados de una prueba de precisión general pueden ocultar disparidades entre grupos. Las métricas deben evaluarse en condiciones comparables al contexto de despliegue y actualizarse cuando cambian las cámaras, la iluminación, el entorno o las características de la población. Una tasa de error aceptable en laboratorio no garantiza que sea aceptable en una decisión policial con efectos sobre libertad o seguridad.

La supervisión externa es especialmente importante cuando los sistemas se utilizan en investigaciones secretas o en tiempo real. Un modelo de control puede combinar autorización previa, auditorías independientes, conservación de registros y acceso posterior a información para personas afectadas, respetando límites necesarios para proteger investigaciones legítimas. La reserva de una investigación no debería convertirse en una exención general frente al escrutinio judicial o administrativo.

8.3.3. Valoración de riesgo en el proceso penal

Los sistemas de valoración de riesgo pueden utilizarse en decisiones de libertad provisional, supervisión comunitaria, clasificación penitenciaria o evaluación de reincidencia. Su uso resulta problemático cuando una probabilidad de grupo se interpreta como un juicio individual de culpabilidad o cuando un resultado opaco condiciona una decisión sobre libertad. El derecho penal exige individualizar la responsabilidad y respetar la presunción de inocencia; un patrón estadístico no prueba que una persona haya cometido un delito.

En la etapa de sentencia, una herramienta puede influir sobre la intensidad de la pena o las condiciones de supervisión. La defensa debería poder conocer que se utilizó un sistema, acceder a sus factores relevantes y cuestionar su fiabilidad y pertinencia. La confidencialidad comercial no debe impedir que el acusado impugne un elemento que puede afectar su libertad. Si la opacidad hace imposible una contradicción adecuada, el tribunal debe considerar si el resultado puede ser utilizado y con qué peso.

State v. Loomis ofrece una ilustración concreta: la Corte Suprema de Wisconsin permitió considerar COMPAS bajo limitaciones, advirtiendo que sus puntajes no podían ser determinantes y que el tribunal debía tener en cuenta las cautelas del instrumento. La decisión no avaló que un algoritmo decida la pena ni resolvió todos los problemas de sesgo, transparencia o evidencia. Su enseñanza más acotada es que la herramienta no reemplaza la motivación judicial y que sus límites

deben ser conocidos por quien decide.

La valoración algorítmica también plantea la diferencia entre riesgo y necesidad de intervención. Un resultado de riesgo elevado no justifica automáticamente una pena más severa; la medida debe tener fundamento legal y relacionarse con los fines legítimos de la decisión. En materia penitenciaria, la predicción puede afectar acceso a programas, clasificación o condiciones de libertad. Las personas privadas de libertad deben contar con mecanismos para conocer y revisar las decisiones que utilizan estos sistemas.

8.3.4. Supervisión, registro y límites al uso

Las autoridades de seguridad deberían mantener un inventario público de sistemas relevantes, con información sobre la finalidad, la autoridad responsable, los tipos de datos y los mecanismos de supervisión. La publicación debe proteger información que legítimamente requiera reserva, pero permitir el escrutinio de los usos y de sus efectos. También deben establecerse protocolos para suspender sistemas cuando presentan fallas, generan impactos discriminatorios o no pueden ser auditados.

La supervisión humana es insuficiente si el funcionario carece de capacidad para cuestionar el resultado o si la puntuación se presenta como científicamente objetiva. Las instituciones deben formar al personal para comprender falsos positivos, tasas de error, límites de predicción y diferencias entre correlación y causalidad. También necesitan reglas para documentar cuándo una recomendación fue aceptada, modificada o descartada.

Las prohibiciones deben ser específicas y ejecutables. Una norma que prohíba genéricamente “el uso perjudicial de IA” puede dejar amplios márgenes de arbitrariedad; una lista demasiado cerrada puede no cubrir nuevas formas de vigilancia o identificación. El diseño jurídico debe

definir finalidades prohibidas, excepciones controladas, autoridades responsables y consecuencias del incumplimiento.

En América Latina, estas decisiones requieren especial atención al contexto institucional, la desigualdad y la capacidad de supervisión. La compra de tecnología importada no elimina la obligación de evaluar su desempeño local ni la compatibilidad con las garantías constitucionales. Una herramienta puede funcionar técnicamente y ser, sin embargo, inadecuada para las condiciones sociales, lingüísticas y normativas del país donde se despliega.

Tabla 8.1

Usos estatales de IA, riesgos y controles mínimos

Ámbito estatal	Uso posible de IA	Riesgo principal	Control mínimo que debe evaluarse
Administración pública	Clasificar solicitudes, priorizar expedientes o apoyar la asignación de prestaciones	Errores que retrasen, denieguen o condicionen el acceso a servicios	Informar sobre el uso del sistema, verificar los datos y permitir revisión por una persona competente
Gestión de servicios públicos	Predecir demanda o distribuir recursos	Asignación desigual por datos incompletos o patrones históricos	Evaluar la calidad de los datos, revisar impactos diferenciados y documentar los criterios de asignación

Ámbito estatal	Uso posible de IA	Riesgo principal	Control mínimo que debe evaluarse
Administración de justicia	Buscar, ordenar o resumir documentos y precedentes	Información incompleta o incorrecta que influya en el análisis jurídico	Verificar fuentes y resultados; mantener la responsabilidad profesional de quien utiliza la herramienta
Apoyo a la función judicial	Facilitar tareas auxiliares o análisis de información	Influencia indebida sobre la independencia y el razonamiento judicial	Usar la herramienta como apoyo, documentar su intervención y preservar la decisión y motivación judiciales
Seguridad y policía	Analizar patrones delictivos o apoyar la planificación de recursos	Vigilancia desproporcionada, sesgos o concentración de controles sobre ciertos grupos	Definir base jurídica, finalidad y límites; supervisar el desempeño y registrar el uso
Proceso penal	Apoyar evaluaciones de riesgo o análisis de evidencia	Tratar una puntuación como prueba concluyente o factor determinante	Examinar las limitaciones, contrastar la información y permitir que las partes cuestionen los elementos relevantes

Nota. *Los controles deben ajustarse al uso, al nivel de riesgo y a las normas aplicables. La tabla no autoriza por sí misma ningún uso de IA. Elaboración propia con base en los temas del capítulo y en las garantías de legalidad, debido proceso, igualdad, transparencia y supervisión humana allí examinadas.*

8.4. Control judicial, transparencia y límites del uso estatal de la IA

8.4.1. Control de legalidad y revisión de decisiones

El control judicial debe poder examinar tanto la decisión final como la forma en que la tecnología influyó en ella. La pregunta no es únicamente si el funcionario firmó el acto, sino si ejerció un juicio independiente y si las razones de la decisión pueden ser verificadas. Cuando el sistema actúa como recomendación, el tribunal puede analizar su peso real; cuando el sistema determina el resultado, el escrutinio debe considerar la base legal de esa automatización y las garantías disponibles.

Una revisión efectiva exige acceso a evidencia suficiente. Puede incluir registros de entrada y salida, la versión del modelo, instrucciones de uso, criterios de evaluación y documentación de las limitaciones. No siempre será necesario divulgar públicamente información comercial sensible, pero la reserva no puede anular el derecho de defensa ni impedir que el juez valore la fiabilidad del sistema. Pueden utilizarse medidas de protección, acceso reservado o peritajes independientes para conciliar intereses legítimos.

El control judicial también debe examinar proporcionalidad. La administración o la policía debe mostrar que la finalidad perseguida es legítima, que el sistema es adecuado para alcanzarla y que no existía una alternativa menos lesiva igualmente eficaz. La proporcionalidad requiere valorar tanto los beneficios alegados como los daños potenciales, incluidos los impactos desiguales y el riesgo de error.

La revisión puede extenderse a patrones, no solo a casos individuales. Si una herramienta produce una tasa inusual de decisiones adversas para un grupo, un tribunal o una autoridad de control puede requerir evidencia sobre su desempeño y sus criterios de uso. La acumulación de reclamaciones puede revelar un problema sistémico que no aparece en un expediente aislado. Para ello, resulta esencial conservar datos y

permitir análisis agregados compatibles con la privacidad.

8.4.2. Transparencia y explicabilidad institucional

La transparencia debe adaptarse a la audiencia y al tipo de decisión. La persona afectada necesita información clara y práctica sobre el rol del sistema, los factores relevantes y la forma de impugnar. El órgano de control necesita documentación más detallada para evaluar riesgos, contratos, auditorías y rendimiento. El público, por su parte, requiere información suficiente para conocer qué sistemas utiliza el Estado y con qué fines.

La transparencia no equivale a publicar código fuente. El acceso al código puede ser útil en ciertas auditorías, pero no siempre permite comprender el resultado ni identificar los impactos. Una explicación institucional puede describir el propósito del sistema, los datos que considera, las variables decisivas, sus limitaciones y las condiciones bajo las cuales un funcionario puede apartarse de sus recomendaciones. Debe permitir una crítica informada, no limitarse a exponer jerga técnica.

La documentación debe ser mantenida durante toda la operación del sistema. Una evaluación inicial pierde valor si el modelo cambia, si se incorporan nuevas bases o si la herramienta comienza a utilizarse para otros fines. Las autoridades deben conservar versiones, incidentes, evaluaciones y decisiones de actualización. Cuando el sistema se retira, también es necesario mantener evidencia durante el tiempo requerido para investigar reclamaciones o revisar decisiones anteriores.

Los registros públicos de sistemas estatales pueden aumentar la rendición de cuentas. Deberían identificar, al menos, la entidad responsable, la finalidad, el nivel de automatización, los grupos potencialmente afectados y los canales de reclamación. Su publicación puede revelar dependencias de proveedores, duplicación de herramientas y usos

que requieren evaluación reforzada. No obstante, el registro debe acompañarse de auditoría y capacidad correctiva; la transparencia formal no compensa la ausencia de supervisión.

8.4.3. Límites constitucionales y derechos fundamentales

El uso estatal de IA está limitado por los derechos fundamentales y por las reglas constitucionales sobre ejercicio del poder. En el sector público, los sistemas no pueden utilizarse para eludir garantías de procedimiento, restringir de forma desproporcionada la privacidad, discriminar en el acceso a servicios o afectar la libertad mediante decisiones carentes de fundamento legal. La automatización tampoco transforma una decisión política en una conclusión neutral: las variables seleccionadas y los umbrales aplicados expresan prioridades institucionales.

La independencia judicial exige que la tecnología apoye, pero no sustituya, la función de juzgar. El debido proceso exige que las partes puedan conocer y controvertir elementos que influyeron en la decisión. La igualdad exige examinar resultados diferenciados. La privacidad requiere limitar datos y finalidades. La libertad de expresión y de reunión puede verse afectada por sistemas de vigilancia o clasificación masiva. Estas garantías se conectan con capítulos anteriores y deben evaluarse junto con el contexto y el efecto concreto.

Los sistemas que afectan a grupos vulnerables requieren salvaguardias reforzadas. En educación, salud, migración, asistencia social, justicia penal o protección de la niñez, los errores pueden ocasionar daños acumulativos y difíciles de revertir. El análisis de riesgo debe considerar quién puede corregir una clasificación errónea, si existe asistencia humana y qué ocurre cuando los datos de una persona son incompletos o no reflejan su situación.

El control constitucional también se relaciona con la delegación de

poder. Las autoridades pueden utilizar tecnología para procesar información, pero no deberían transferir a proveedores privados la decisión normativa sobre criterios de elegibilidad, prioridad o coerción. Las decisiones que implican ponderar derechos, interpretar normas o determinar consecuencias deben permanecer bajo responsabilidad de órganos legalmente competentes y estar sujetas a revisión.

8.4.4. Gobernanza y rendición de cuentas estatal

La gobernanza eficaz requiere una distribución clara de funciones entre quienes proponen, adquieren, desarrollan, operan y supervisan sistemas. Debe existir una instancia capaz de detener el uso cuando los riesgos no se controlan y una autoridad independiente que pueda investigar incumplimientos. La entidad responsable debe disponer de personal jurídico, técnico y de política pública, así como de procedimientos para escuchar a las personas afectadas.

El Consejo Nacional de Justicia de Brasil adoptó la Resolución n.º 615/2025 para regular el desarrollo, la utilización y la gobernanza de soluciones de IA en el Poder Judicial. Entre sus componentes constan la auditoría, el monitoreo, la clasificación por riesgo y el registro de sistemas en Sinapses. Este marco sectorial constituye un ejemplo de gobernanza judicial institucionalizada; su alcance se limita al Poder Judicial brasileño y no debe generalizarse como régimen nacional para todos los usos de IA.

En la Unión Europea, el Reglamento de IA incorpora obligaciones para sistemas de alto riesgo, incluidos determinados usos en servicios esenciales, aplicación de la ley y administración de justicia. Su clasificación no reemplaza las reglas procesales nacionales ni el control de derechos fundamentales. Los tribunales y autoridades deben aplicar el Reglamento junto con las normas constitucionales, administrativas, de protección de datos y sectoriales que correspondan.

La rendición de cuentas requiere mecanismos internos y externos. Internamente, las entidades deben conservar registros, asignar responsables y revisar incidentes. Externamente, los tribunales, órganos de auditoría, autoridades de datos y organismos de derechos humanos deben poder acceder a información pertinente y ordenar medidas dentro de sus competencias. La coordinación evita que una decisión automatizada quede fuera de control porque cada autoridad considera que corresponde a otra.

En síntesis, la IA puede mejorar la gestión pública y judicial cuando complementa capacidades humanas, amplía el acceso y reduce tareas repetitivas. Su legitimidad depende de que el Estado no delegue su deber de justificar decisiones, preserve la independencia de quienes juzgan y garantice que las personas puedan impugnar resultados que afectan sus derechos. El principio rector es que la tecnología sirva a la justicia y a la administración sometidas al derecho, y no que el derecho se adapte a resultados que nadie puede explicar ni revisar.

Figura 8.1.

Garantías para una decisión pública asistida por IA



Nota: *Elaboración propia. La decisión y la responsabilidad institucional siguen correspondiendo a la autoridad competente.*

Capítulo 9.

Propiedad intelectual, contratación y actividad empresarial con IA



Introducción

La inteligencia artificial transforma la creación, la explotación y la contratación de bienes y servicios. Una obra puede combinar aportes humanos y resultados generados por un modelo; el entrenamiento puede involucrar reproducciones de obras protegidas; y una operación contractual puede ejecutarse mediante código sin que todas sus condiciones sean visibles para quienes participan. Al mismo tiempo, los modelos, datos, capacidades de cómputo e infraestructuras digitales pueden concentrarse en un número reducido de empresas.

El análisis jurídico requiere distinguir cuestiones que suelen confundirse: quién puede ser autor; quién posee los derechos patrimoniales sobre una obra asistida por IA; cuándo el entrenamiento requiere autorización; qué efectos tiene un contrato automatizado; y cómo se protegen la competencia y los activos intangibles. Este capítulo examina esos problemas desde el derecho comparado, con atención a las normas de la Comunidad Andina, la Unión Europea y Estados Unidos, sin presentar una solución nacional como regla universal.

9.1. Autoría, titularidad y protección de obras y creaciones generadas por IA

9.1.1. La autoría y el aporte humano

El derecho de autor protege expresiones originales, pero los sistemas jurídicos no coinciden completamente sobre quién puede ser autor ni sobre el grado de intervención humana necesario. En los países miembros de la Comunidad Andina, la Decisión 351 define al autor como la persona física que realiza la creación intelectual. Esta definición es relevante para Ecuador, Colombia, Perú y Bolivia: bajo ese régimen común, una IA no aparece como autora en el sentido jurídico ordinario.

De esa premisa no se sigue que toda obra producida con asistencia algorítmica carezca de protección. La cuestión consiste en identificar qué elementos expresivos proceden de una contribución humana suficientemente original. Una persona puede seleccionar, ordenar, corregir, editar o integrar resultados del sistema de modo que la obra final exprese decisiones creativas propias. En tal caso, los componentes humanos pueden ser protegibles, aunque el resultado contenga material generado automáticamente que no cumpla los requisitos de autoría del ordenamiento aplicable.

La intensidad y naturaleza del aporte importan más que el simple hecho de utilizar una herramienta. Quien usa IA para revisar ortografía, traducir un borrador o proponer alternativas no necesariamente deja de ser autor de la obra que concibió y redactó. En cambio, una instrucción breve que produce un texto o una imagen sin que la persona determine de forma significativa sus elementos expresivos puede no bastar para atribuirle autoría sobre todo el resultado. La línea debe trazarse a partir del proceso creativo, de la legislación nacional y de la prueba disponible.

En Estados Unidos, la Oficina de Derecho de Autor concluyó en su informe de 2025 que el material generado exclusivamente por

IA no recibe protección bajo las reglas vigentes, mientras que las contribuciones humanas perceptibles, la selección o disposición creativa y las modificaciones originales pueden ser protegibles. La Oficina añadió que la mera provisión de instrucciones o *prompts*, por sí sola, no suele establecer control suficiente sobre los elementos expresivos. Se trata de la interpretación administrativa de Estados Unidos, no de una regla internacional vinculante.

Este enfoque permite reconocer una obra híbrida: una parte puede ser protegible y otra quedar fuera de la exclusividad, según el grado de autoría humana. En un libro, por ejemplo, la selección de fuentes, la estructura argumentativa, la redacción y la revisión crítica pueden expresar decisiones humanas, mientras que ciertos pasajes generados por un modelo podrían no estar protegidos de manera independiente. El análisis no se resuelve con una declaración general de que "se utilizó IA"; requiere identificar qué aportes realizó cada participante.

9.1.2. Obra asistida, obra generada y obra colectiva

Conviene distinguir entre obra asistida por IA y obra generada por IA. En la primera, la persona dirige un proceso creativo, utiliza la herramienta como apoyo y conserva control sobre decisiones expresivas sustanciales. En la segunda, el sistema produce una salida con intervención humana limitada. Esta distinción no siempre es nítida, porque los usuarios pueden iterar, descartar y modificar múltiples resultados; sin embargo, resulta útil para analizar quién creó los elementos protegibles.

La intervención puede aparecer en distintas fases: elección del tema y propósito; elaboración de instrucciones; selección entre resultados; corrección de contenido; combinación con materiales propios; y edición final. No todas esas actividades tienen el mismo peso creativo. La decisión de aceptar una salida sin cambios puede revelar un acto de selección, pero no necesariamente control sobre la forma de expresión generada. A la inversa, una intervención extensa de edición puede

transformar el resultado hasta producir una expresión propia.

La atribución también debe contemplar a quienes intervienen en el proyecto. Una persona puede concebir y editar; otra puede seleccionar materiales y realizar la composición visual; una entidad puede encargar la obra y adquirir derechos patrimoniales conforme a la legislación nacional y al contrato. La titularidad patrimonial no debe confundirse con la autoría: una empresa puede ser titular derivada de derechos sobre una obra humana sin ser autora de ella. La Decisión 351 diferencia la persona autora de la posibilidad de que personas naturales o jurídicas ejerzan derechos patrimoniales conforme a la ley aplicable y a acuerdos válidos.

En una obra colectiva o por encargo, la institución debería conservar evidencia de quién realizó las contribuciones humanas, qué herramienta se utilizó y qué términos regían su uso. Esta documentación puede ser necesaria para acreditar titularidad, respetar derechos morales, responder a una reclamación o negociar licencias. Una cláusula que declare que "todo lo generado pertenece al cliente" no puede crear derechos de autor sobre material que el derecho aplicable excluya de protección, aunque sí puede distribuir entre las partes derechos contractuales sobre servicios, archivos, entregables o información confidencial.

9.1.3. Registro, transparencia y gestión de derechos

La gestión editorial y empresarial debe diferenciar tres capas: materiales protegidos que una persona aporta; contenido generado o transformado mediante IA; y aportes humanos originales realizados durante la selección, edición o integración. Esta distinción ayuda a evitar declaraciones de autoría imprecisas y a aclarar qué derechos pueden licenciarse. También facilita verificar que no se han incorporado obras de terceros de manera incompatible con sus condiciones de uso.

El registro puede requerir declarar la presencia de material generado por IA o excluirlo de la pretensión de autoría, dependiendo de la jurisdicción y del trámite. En Estados Unidos, la Oficina de Derecho de Autor exige identificar el material generado por IA en solicitudes de registro cuando corresponda y describir la contribución humana. La práctica de registro no determina de forma definitiva la validez de todos los derechos ni reemplaza el análisis judicial si existe una controversia.

En la Comunidad Andina, la definición de autor como persona física orienta la titularidad originaria, pero la Decisión 351 reconoce derechos morales y patrimoniales con reglas propias. Por ello, conviene evitar afirmar que una obra generada sin intervención humana queda automáticamente en dominio público en todos los países. La conclusión depende de si el ordenamiento protege la expresión, reconoce alguna categoría especial o aplica reglas sobre titularidad y explotación distintas. En ausencia de una regla específica, la exclusividad de derecho de autor no debe presumirse para resultados que no satisfacen el requisito de autoría humana.

La transparencia sobre uso de IA puede ser relevante para contratos editoriales, concursos, investigación y encargos creativos. No existe una obligación general idéntica para todas las obras y jurisdicciones; puede haber obligaciones derivadas de licencias, términos de plataformas, normas sectoriales o reglas de integridad académica. La divulgación debe ser proporcional al contexto y no sustituir el análisis de autoría. Indicar que una obra fue "creada con IA" no resuelve quién posee los derechos ni si el uso de los materiales de entrada fue lícito.

9.1.4. Propiedad industrial y activos intangibles

La IA interviene también en la innovación industrial. Puede apoyar la búsqueda de soluciones técnicas, el diseño de productos, el análisis de mercados y la generación de prototipos. Sin embargo, las patentes siguen exigiendo cumplir requisitos legales de invención, novedad, actividad inventiva y aplicación industrial, según cada jurisdicción. La

herramienta puede asistir en el proceso, pero no es automáticamente reconocida como inventora humana. El equipo y la empresa deben identificar a las personas que concibieron la solución conforme al estándar aplicable y mantener registros de sus contribuciones.

Las marcas plantean un análisis diferente. Un nombre, logotipo o signo generado con IA puede ser susceptible de registro si cumple los requisitos de distintividad y no infringe derechos anteriores. El uso de modelos para producir imágenes comerciales requiere verificar licencias, similitudes con signos de terceros y restricciones contractuales. El hecho de que un resultado sea generado por un sistema no garantiza que esté libre de conflictos marcarios o de derechos de imagen.

Los secretos empresariales pueden comprender parámetros, procesos, bases de datos, instrucciones, evaluaciones o información comercial no divulgada, siempre que se cumplan los requisitos legales de secreto y medidas razonables de protección. Los datos no se convierten automáticamente en secreto comercial por ser valiosos; su protección depende de que no sean generalmente conocidos, tengan valor por su carácter secreto y estén sujetos a medidas de confidencialidad. La empresa debe evitar divulgar información reservada al cargarla en herramientas públicas cuyos términos permitan retenerla o reutilizarla.

Estos activos pueden coexistir. Un mismo sistema puede estar protegido mediante derechos de autor sobre el software, patentes sobre una invención técnica, marcas sobre sus signos y secreto empresarial sobre ciertos datos o parámetros. La elección de una estrategia de protección debe considerar divulgación, duración, facilidad de copia, posibilidad de ingeniería inversa y necesidades de interoperabilidad. No toda innovación debe patentarse ni toda información valiosa debe mantenerse secreta.

9.2. Datos de entrenamiento, derechos de autor y uso legítimo

9.2.1. Reproducción, minería de textos y datos

El entrenamiento de modelos generativos puede implicar la recopilación, copia, almacenamiento, transformación y análisis de grandes cantidades de textos, imágenes, música, código y otros materiales. Algunos están protegidos por derecho de autor; otros se encuentran en dominio público, están licenciados o quedan comprendidos en excepciones legales. La cuestión jurídica no es simplemente si una obra estaba disponible en internet, sino qué actos se realizaron, con qué finalidad, en qué jurisdicción y bajo qué base legal o contractual.

La minería de textos y datos (TDM) es un conjunto de técnicas automatizadas para analizar material con el fin de extraer patrones, tendencias o información. Puede incluir reproducciones técnicas, aunque el resultado no sustituya ni reproduzca expresivamente la obra original. La licitud de esas reproducciones depende de la ley aplicable y de sus excepciones, limitaciones y condiciones. No debe suponerse que toda minería para entrenamiento generativo es automáticamente lícita, ni que toda copia temporal constituye necesariamente infracción.

La Directiva (UE) 2019/790 incorpora dos excepciones relevantes de minería de textos y datos. El artículo 3 contempla minería para investigación científica por determinados organismos e instituciones patrimoniales, bajo condiciones de acceso lícito y conservación segura. El artículo 4 establece una excepción más amplia, incluida la minería con fines comerciales, pero permite que los titulares reserven expresamente sus derechos, en particular mediante medios legibles por máquina cuando el contenido está disponible públicamente en línea. Estas reglas son propias de la Unión Europea y no se aplican automáticamente en Ecuador ni en otros países.

La existencia de una excepción no elimina otras obligaciones. Puede ser necesario respetar términos de acceso, protección de bases de datos,

medidas tecnológicas, derechos conexos o normas de protección de datos personales. También importa si el sistema almacena fragmentos extensos o si reproduce obras identificables en sus salidas. El análisis del entrenamiento y el de la generación deben distinguirse: una utilización puede ser lícita en la fase de análisis y, aun así, una salida concreta podría infringir derechos si reproduce una expresión protegida sin autorización o excepción aplicable.

9.2.2. Opt-out, licencias y transparencia

El régimen europeo de minería con fines generales permite a los titulares reservar derechos. En la práctica, esto plantea cuestiones sobre cómo se expresa una reserva, cómo la detectan los sistemas de recopilación y cómo se prueba que el proveedor la respetó. Los mecanismos técnicos de exclusión pueden ayudar, pero su eficacia depende de estándares interoperables, accesibilidad y prácticas verificables.

Las licencias constituyen una alternativa para obtener autorización, fijar condiciones de uso y, en su caso, remunerar a titulares. Pueden negociarse individualmente, a través de entidades de gestión colectiva o mediante licencias colectivas ampliadas cuando la legislación nacional lo permita. La licencia debe aclarar qué obras cubre, qué usos se autorizan, qué territorios y periodos comprende, cómo se calculan los pagos, si se permite el entrenamiento comercial y cómo se gestionan las salidas que reproduzcan material protegido.

La transparencia ayuda a que los titulares y las autoridades evalúen el cumplimiento. El Reglamento de IA de la Unión Europea exige a los proveedores de modelos de propósito general adoptar una política para cumplir la normativa de derecho de autor aplicable y elaborar un resumen suficientemente detallado del contenido utilizado para entrenamiento, conforme a las disposiciones del Reglamento. Esta obligación no significa que el resumen deba revelar cada archivo ni que resuelva por sí solo la licitud de cada reproducción. Sí establece un deber regulatorio de transparencia y de política de cumplimiento para

los proveedores comprendidos en su ámbito.

Las obligaciones europeas se complementan con los instrumentos de orientación para proveedores de modelos de propósito general. Sin embargo, el cumplimiento de una política de derechos de autor y la publicación de un resumen no constituyen automáticamente autorización de cada titular. Tampoco eliminan la posibilidad de que una obra concreta haya sido utilizada en condiciones incompatibles con la ley nacional. La evaluación continua dependiendo del marco aplicable, el tipo de obra, la finalidad, el acceso, las excepciones y los efectos de la utilización.

9.2.3. Uso legítimo y enfoques comparados

En Estados Unidos, la doctrina de *fair use* permite evaluar usos no autorizados mediante un análisis contextual de factores legales, entre ellos la finalidad del uso, la naturaleza de la obra, la cantidad utilizada y el efecto sobre el mercado potencial. El análisis no produce una respuesta automática para todo entrenamiento de modelos. La Oficina de Derecho de Autor estadounidense, en su informe de 2025 sobre entrenamiento generativo, examinó distintos usos y sostuvo que las cuestiones dependen de hechos concretos; el informe es un documento administrativo de análisis y no una decisión judicial vinculante para todos los casos.

La comparación muestra por qué no debe usarse el término *fair use* como si describiera una excepción universal. La Unión Europea dispone de reglas legislativas específicas para minería de textos y datos; Estados Unidos aplica una doctrina abierta de uso legítimo; los países andinos cuentan con su propio régimen de derechos exclusivos y excepciones, sin que la Decisión 351 contenga una excepción general idéntica a la estadounidense para cualquier entrenamiento comercial. La licitud debe determinarse con referencia al país pertinente y a las normas nacionales y comunitarias aplicables.

La territorialidad complica aún más el análisis. Una empresa puede recopilar obras en un país, procesarlas en servidores de otro y ofrecer el modelo a usuarios de múltiples jurisdicciones. La ubicación del servidor no resuelve necesariamente qué ley se aplica, pues pueden ser relevantes el lugar de la reproducción, el mercado dirigido, el lugar del daño y las reglas de conflicto de leyes. Las empresas necesitan mapear sus operaciones y no asumir que la licencia o excepción válida en una jurisdicción cubre todas las demás.

La incertidumbre normativa no justifica ni la extracción indiscriminada ni una prohibición absoluta del análisis computacional. Un marco equilibrado requiere reglas claras sobre acceso, reservas de derechos, transparencia, licenciamiento y protección de investigación. También debe considerar los intereses de autores y titulares, la innovación de empresas emergentes, la investigación científica y la diversidad cultural y lingüística.

9.2.4. Datos públicos, bases de datos y privacidad

Que un material sea accesible públicamente no significa que carezca de protección. Una fotografía publicada en una red social puede seguir protegida por derecho de autor; una base de datos puede incorporar derechos sobre su estructura o sobre la inversión realizada; y la información personal incluida en una página pública continúa sujeta a normas de privacidad. La disponibilidad técnica no equivale a una licencia abierta ni a una autorización general para entrenamiento.

El análisis debe separar los derechos involucrados. La extracción de una obra puede afectar derechos de autor; la recopilación de datos personales puede activar obligaciones de protección de datos; y el uso de contenido de una plataforma puede contravenir términos contractuales. Un conjunto de datos puede incluir obras en dominio público, material bajo licencias abiertas y material protegido sin permiso. La trazabilidad de las fuentes y de las licencias reduce el riesgo y facilita corregir errores.

El uso de datos personales para entrenamiento puede requerir una base legal, información a los titulares y evaluación de compatibilidad de finalidad, según el régimen aplicable. El hecho de que el modelo no reproduzca directamente un registro individual no resuelve por sí solo la legalidad del tratamiento durante la fase de entrenamiento. A la vez, el análisis de privacidad no sustituye el examen de derechos de autor: son preguntas jurídicas distintas que pueden concurrir sobre los mismos datos.

Las empresas que crean modelos deberían documentar procedencia, permisos, reservas de derechos, filtros aplicados y mecanismos de exclusión. Para los titulares, la evidencia puede incluir marcas de agua, registros de publicación, avisos de reserva, contratos y pruebas de que el modelo genera pasajes similares. La transparencia técnica es necesaria para que una controversia pueda resolverse sobre hechos verificables en lugar de afirmaciones generales sobre los datos de entrenamiento.

Tabla 9.1

Preguntas jurídicas en el ciclo empresarial de IA

Fase del ciclo empresarial	Aspectos jurídicos que deben revisarse	Pregunta de análisis
Obtención y preparación de datos	Procedencia, derechos de autor, datos personales, licencias y restricciones de uso	¿La organización cuenta con una base jurídica y permisos suficientes para recopilar y utilizar los datos?
Entrenamiento o configuración	Minería de textos y datos, reservas de derechos, condiciones de acceso y documentación	¿Se han revisado las excepciones, licencias y restricciones aplicables en la jurisdicción correspondiente?

Fase del ciclo empresarial	Aspectos jurídicos que deben revisarse	Pregunta de análisis
Generación y uso de resultados	Aporte humano, titularidad, semejanzas con obras existentes y derechos de terceros	¿Qué aportes humanos pueden identificarse y qué derechos podrían verse afectados por el resultado?
Contratación de proveedores	Finalidad autorizada, datos, actualizaciones, seguridad, auditoría y distribución de responsabilidades	¿El contrato permite conocer, supervisar y cuestionar el funcionamiento del servicio, y define cómo se atenderán los incidentes?
Despliegue en la empresa	Usos previstos, supervisión, cumplimiento normativo y mecanismos de reclamación	¿La organización evalúa los riesgos del sistema en el contexto real y asigna responsables para controlarlo?
Competencia y continuidad del servicio	Dependencia de proveedores, interoperabilidad, portabilidad y costos de salida	¿La empresa puede cambiar de proveedor o recuperar sus datos y registros sin obstáculos desproporcionados?

Nota. *La respuesta depende de la legislación aplicable, de los términos contractuales y de las funciones que desempeñe cada actor. Elaboración propia con base en los temas del capítulo.*

9.3. Contratos inteligentes, contratación tecnológica y cumplimiento normativo

9.3.1. Contratación automatizada y validez jurídica

La contratación tecnológica incluye contratos para desarrollar modelos, licenciar software, adquirir servicios en la nube, obtener datos, integrar sistemas y desplegar soluciones de IA. La relación puede involucrar proveedores de modelos base, intermediarios, integradores y clientes finales. Los contratos deben determinar qué servicio se promete, quién conserva control sobre los datos y las actualizaciones, qué garantías

se ofrecen y cómo se gestionan errores, incidentes, auditorías y terminación.

La automatización no elimina los requisitos generales del contrato. Deben existir consentimiento, capacidad, objeto y demás elementos que establezca la ley aplicable; además, pueden ser relevantes las formalidades, la protección del consumidor y las normas de contratación pública. Un sistema puede aceptar una oferta o ejecutar una operación en nombre de una organización, pero la atribución jurídica de esa conducta depende de las reglas sobre representación, autorización y comunicaciones electrónicas.

La Comisión de las Naciones Unidas para el Derecho Mercantil Internacional (CNUDMI) aprobó en 2024 la Ley Modelo sobre Contratación Automatizada, destinada a facilitar operaciones contractuales mediante IA, sistemas automatizados y contratos inteligentes. La Ley Modelo reconoce el uso de código informático e información dinámica para formar y ejecutar operaciones, pero es un texto legislativo modelo: no es un tratado ni se convierte automáticamente en norma vigente para cada país. Su aplicación requiere que el Estado la incorpore o adapte a su derecho interno.

La contratación electrónica también se apoya en principios de no discriminación por el uso de medios electrónicos, equivalencia funcional y neutralidad tecnológica. Estos principios permiten que un contrato no pierda eficacia únicamente por estar formado o expresado electrónicamente. No resuelven, sin embargo, todos los problemas de automatización: pueden persistir dudas sobre la atribución de una operación, el error, la representación, la revocación o la modificación de condiciones.

9.3.2. Contratos inteligentes y código informático

Un contrato inteligente suele referirse a código que ejecuta

automáticamente ciertas instrucciones, con frecuencia en una cadena de bloques. El código puede transferir activos, liberar pagos, registrar condiciones o activar funciones al verificarse un evento. No todo contrato inteligente es un contrato jurídico completo: puede ser un mecanismo de ejecución de un acuerdo expresado en lenguaje natural, un componente técnico de una relación más amplia o un programa que implementa una regla sin abarcar todas las obligaciones y remedios.

La CNUDMI aclara que la Ley Modelo de Contratación Automatizada reconoce el uso jurídico de código informático, pero no elimina la necesidad de interpretar las reglas contractuales aplicables. Las partes pueden acordar que el código ejecute una operación, pero el sistema jurídico aún debe resolver qué sucede ante error, fraude, fuerza mayor, vulnerabilidad o ejecución incompatible con el contrato. La irreversibilidad técnica de una transacción no significa que sea jurídicamente imposible solicitar restitución o reparación.

La redacción contractual debe conectar el código con el acuerdo jurídico. Conviene definir qué versión del programa controla, cómo se incorporan las condiciones, qué idioma prevalece en caso de discrepancia, quién puede actualizar o detener la operación, qué procedimientos aplican ante fallos y qué ley y jurisdicción resolverán controversias. También debería contemplar oráculos, es decir, fuentes externas de información que activan el código, pues el sistema puede funcionar correctamente y ejecutar una operación basada en datos incorrectos.

En el caso de IA, la automatización puede ser dinámica: el sistema genera una recomendación, activa un flujo, negocia términos o toma acciones conforme a instrucciones preconfiguradas. Los contratos deben distinguir entre decisiones autorizadas automáticamente y actos que requieren aprobación humana. Si el sistema se aparta de parámetros o utiliza datos erróneos, las partes necesitan conocer cómo se suspende su operación y quién asume las consecuencias.

9.3.3. Contratación de IA: datos, actualizaciones y auditoría

Los contratos de IA deberían describir con precisión el uso autorizado y las limitaciones del sistema. Una cláusula genérica de "uso para fines comerciales" puede ser insuficiente si las partes necesitan establecer si los datos del cliente se utilizarán para entrenar modelos, si se retendrán después de terminar el servicio o si podrán compartirse con subcontratistas. La finalidad, la seguridad, la conservación, la portabilidad y la eliminación de datos deben ser explícitas.

Las garantías de desempeño requieren cautela. Un proveedor puede garantizar disponibilidad, tiempos de respuesta o ciertos niveles de servicio; es más complejo asegurar una tasa constante de precisión en contextos que cambian con los datos y la población. El contrato puede fijar métricas, escenarios de prueba, procedimientos de validación y obligaciones de actualización, pero no debe presentar probabilidades estadísticas como garantías absolutas de resultado.

La auditoría debe ser compatible con la protección de secretos empresariales, sin volverse ilusoria. Puede preverse acceso de auditores independientes a documentación y registros; pruebas sobre muestras representativas; notificación de incidentes; y cooperación para responder a reclamaciones. Si el cliente no puede conocer limitaciones esenciales ni demostrar cómo se produjo una decisión, la distribución contractual del riesgo puede dejar a los afectados sin remedio práctico.

La contratación pública requiere condiciones adicionales. Las entidades estatales deben conservar autoridad sobre la decisión, tener acceso a información suficiente para motivarla y evitar dependencia tecnológica que impida cambiar de proveedor. La posibilidad de exportar datos, migrar modelos, recuperar registros y mantener la continuidad del servicio debe analizarse antes de contratar. También resulta necesario prever controles de ciberseguridad, interoperabilidad y cumplimiento de las normas administrativas, de protección de datos y de derechos

fundamentales.

El cumplimiento normativo debe incorporarse a todo el ciclo contractual. Las partes pueden asignar quién realiza evaluaciones de impacto, quién mantiene registros, quién comunica incidentes y quién responde a requerimientos regulatorios. La cláusula no elimina las obligaciones legales frente a autoridades o terceros; ordena la cooperación y, eventualmente, la distribución interna de costos. Una asignación contractual no puede convertir al proveedor en autoridad pública ni exonerar al organismo que toma una decisión bajo su competencia.

9.3.4. Propiedad intelectual y límites de las cláusulas

Los contratos tecnológicos suelen regular la titularidad de software, documentación, datos, mejoras, modelos adaptados y resultados generados. Es importante diferenciar materiales preexistentes del proveedor, materiales del cliente, desarrollos conjuntos y resultados de IA cuya protección legal puede ser incierta. La cláusula debe evitar atribuir derechos exclusivos sobre elementos que no sean protegibles bajo la ley aplicable, sin dejar de regular el acceso y uso contractual de los entregables.

En la Comunidad Andina, la Decisión 351 dispone que las cesiones y licencias de uso se rigen por el contrato y por las legislaciones nacionales; también establece que la transferencia se limita a las formas de explotación expresadas en el acuerdo. En consecuencia, el alcance de una licencia debe ser específico respecto de reproducción, adaptación, distribución, comunicación pública, territorio, plazo y modalidad de explotación.

Las cláusulas de confidencialidad y secreto empresarial pueden proteger información valiosa, pero no deben impedir el cumplimiento de obligaciones legales ni la investigación de vulneraciones de derechos.

Los acuerdos de licencia deben también establecer si el cliente puede utilizar resultados en publicaciones, entrenar sistemas adicionales o sublicenciar entregables. La ausencia de claridad puede producir disputas sobre materiales híbridos, sobre todo cuando el cliente aporta datos y el proveedor integra modelos propios.

Por último, los contratos no deben emplearse para trasladar todo el riesgo al usuario final mediante términos incomprensibles o no negociables. En relaciones de consumo, empleo o servicios esenciales, pueden aplicarse reglas imperativas sobre transparencia, cláusulas abusivas y protección de datos. La validez de una cláusula depende del ordenamiento y del tipo de relación; la aceptación mediante un clic no demuestra necesariamente que la persona comprendió el alcance de una autorización amplia.

9.4. Competencia, mercados digitales y gobernanza corporativa de la IA

9.4.1. Concentración en infraestructura, datos y modelos

La competencia en IA se desarrolla en varias capas: semiconductores y capacidad de cómputo; servicios de nube; conjuntos de datos; modelos fundacionales; herramientas de desarrollo; y aplicaciones para usuarios y empresas. Una empresa que controle recursos clave en varias capas puede obtener ventajas para mejorar sus modelos y distribuirlos, además de influir en las condiciones de acceso de competidores. La integración vertical no es ilícita por sí misma, pero puede plantear riesgos cuando cierra mercados, encarece insumos o limita la interoperabilidad.

La Comisión Federal de Comercio de Estados Unidos (FTC) examinó asociaciones e inversiones entre proveedores de servicios en la nube y desarrolladores de IA generativa. Su informe de personal analizó posibles efectos sobre acceso a capacidad de cómputo, dependencia, costos de cambio e información comercial sensible. Es un estudio de mercado, no una decisión que declare ilícita cada asociación

examinada. Su valor consiste en identificar mecanismos competitivos que las autoridades pueden investigar.

La concentración puede surgir por costos fijos elevados, economías de escala, efectos de red, acceso privilegiado a datos o dependencia de infraestructuras. También puede reforzarse mediante acuerdos de exclusividad, inversiones con derechos de acceso preferente o integración de servicios en plataformas dominantes. Para determinar si existe una infracción, la autoridad debe definir mercado relevante, poder de mercado, conducta y efectos, conforme a las reglas de competencia aplicables. No basta con observar que una empresa es grande o que un modelo tiene muchos usuarios.

La política de competencia debe considerar tanto riesgos como eficiencias. Una alianza puede facilitar inversión, investigación e innovación; también puede excluir a competidores emergentes del acceso a infraestructura o talento. El análisis requiere observar contratos, condiciones de acceso, interoperabilidad, costos de cambio y capacidad de los rivales para alcanzar clientes. Las autoridades necesitan conocimiento técnico suficiente para examinar la cadena de valor y evitar tratar todos los modelos, proveedores y mercados como si fueran iguales.

9.4.2. Plataformas, datos y competencia desleal

Las plataformas digitales pueden integrar IA para clasificar resultados, recomendar productos, ordenar publicidad, resumir contenidos o prestar servicios a empresas usuarias. Si una plataforma favorece sistemáticamente sus propios servicios, combina datos de forma anticompetitiva o dificulta la portabilidad, puede afectar la competencia y las opciones de consumidores y empresas. El análisis debe distinguir el uso legítimo de datos para mejorar servicios de prácticas que excluyen o discriminan a rivales.

En la Unión Europea, el Reglamento de Mercados Digitales (DMA) establece obligaciones para determinados guardianes de acceso en servicios básicos de plataforma. No es una regulación general de IA: su aplicación depende de la designación de la empresa y del servicio comprendido. Sin embargo, algunas prácticas relacionadas con datos, interoperabilidad y acceso pueden incidir en mercados donde se incorporan herramientas de IA.

El acceso a datos puede ser una fuente de ventaja, aunque no todos los datos son intercambiables ni deben compartirse sin límites. La protección de datos personales, los derechos de autor y los secretos empresariales siguen siendo relevantes. La política de competencia puede exigir examinar si una práctica excluye rivales, pero no puede ignorar otros derechos ni imponer una apertura indiscriminada que incremente riesgos de privacidad o seguridad.

Las autoridades también pueden estudiar adquisiciones, acuerdos de inversión y asociaciones antes de que consoliden una dependencia. Un competidor potencial puede ser adquirido antes de alcanzar una posición significativa, y una alianza puede vincular acceso a nube, distribución y desarrollo de modelos. La evaluación debe ser caso por caso; no toda inversión en una empresa de IA supone una concentración ilícita, pero tampoco todo acuerdo comercial está fuera del escrutinio por llevar la etiqueta de "innovación".

9.4.3. Cumplimiento corporativo y gobernanza interna

La gobernanza corporativa de IA comprende las estructuras mediante las cuales una organización decide qué sistemas desarrollar o adquirir, cómo supervisarlos y quién responde por sus impactos. No se limita a un comité de ética ni a una política pública de principios. Incluye responsabilidades de directorio y gerencia, controles de riesgo, gestión de datos, ciberseguridad, documentación, capacitación y canales de denuncia.

Una gobernanza eficaz establece inventarios de sistemas, clasifica usos por riesgo y asigna responsables de negocio, técnicos, jurídicos y de cumplimiento. Antes de desplegar una herramienta, la organización debe evaluar finalidad, datos, usuarios afectados, proveedores, resultados posibles y vías de reclamación. Durante la operación, debe monitorear cambios y registrar incidentes. Cuando el sistema ya no cumple condiciones de seguridad o legalidad, debe poder suspenderse.

La empresa debe también verificar la procedencia de datos y derechos asociados. El uso de datos de entrenamiento puede implicar licencias, reservas, privacidad y términos contractuales. La generación de resultados puede producir semejanzas no deseadas, información falsa o usos incompatibles con marcas y derechos de terceros. Los controles previos y posteriores —revisión de muestras, filtros y protocolos de escalamiento— reducen riesgos, aunque no garantizan ausencia total de infracciones.

El cumplimiento debe integrarse en las decisiones de negocio. Un programa que depende exclusivamente del equipo técnico puede carecer de autoridad para limitar lanzamientos; uno que depende únicamente del área jurídica puede no detectar fallas en datos o desempeño. La cooperación interfuncional permite evaluar riesgos desde perspectivas distintas y documentar por qué se adoptó una decisión. La alta dirección debe recibir información comprensible sobre riesgos significativos, incidentes y medidas correctivas.

9.4.4. Innovación responsable y creación de valor

La propiedad intelectual y la competencia no son obstáculos externos a la innovación: determinan quién puede crear, acceder, licenciar y beneficiarse de los sistemas. Un marco previsible puede facilitar inversión y colaboración, mientras que la concentración excesiva puede reducir opciones y elevar barreras de entrada. Del mismo modo, la protección de autores debe coexistir con investigación y usos permitidos por la ley; el equilibrio requiere reglas claras y mecanismos

de licencia y transparencia.

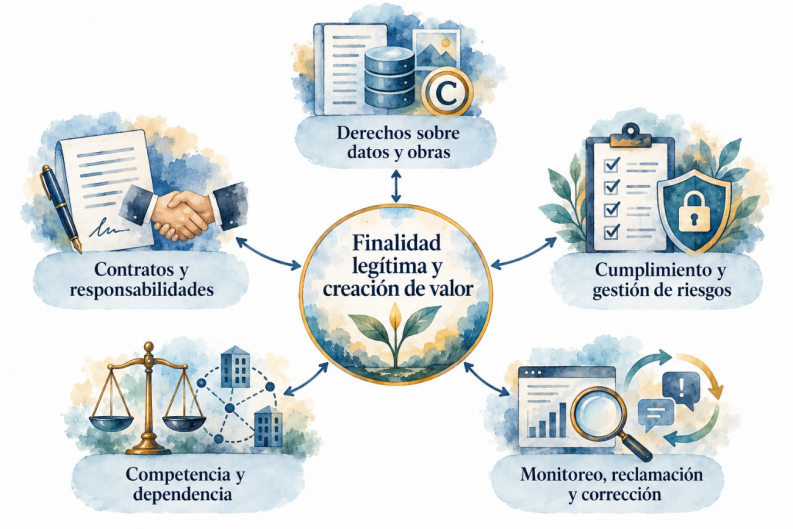
Las empresas emergentes y las instituciones de investigación suelen depender de servicios de nube y modelos de terceros. Contratos con costos de salida elevados, restricciones sobre portabilidad o falta de acceso a registros pueden crear dependencia. Las organizaciones deben considerar estrategias de interoperabilidad, disponibilidad de proveedores alternativos y posibilidades de migración. Las políticas públicas pueden promover competencia mediante estándares abiertos, compras públicas competitivas y apoyo a investigación, sin prescindir de requisitos de seguridad y protección de derechos.

La responsabilidad corporativa incluye evaluar efectos sobre trabajadores, consumidores y comunidades. Sistemas que automatizan contratación, crédito, atención o vigilancia pueden cambiar la distribución de oportunidades y riesgos. La gobernanza debe permitir que las personas afectadas cuestionen decisiones y que la organización investigue patrones, no solo incidentes aislados. El cumplimiento no se agota en evitar sanciones: también preserva confianza, continuidad operativa y legitimidad social.

En definitiva, la actividad empresarial con IA exige coordinar derechos de autor, propiedad industrial, secretos empresariales, contratación, competencia y protección de datos. Ningún instrumento resuelve por sí solo la totalidad de los problemas. La autoría requiere identificar aportes humanos; el entrenamiento exige revisar derechos y excepciones aplicables; los contratos deben asignar responsabilidades y permitir control; y la gobernanza competitiva debe observar tanto la innovación como el poder de mercado. Esa articulación permite que la IA genere valor económico sin convertir la incertidumbre jurídica en una ventaja para quienes controlan datos, infraestructura y canales de distribución.

Figura 9.1.

Gobernanza empresarial de IA: derechos, contratos y controles



Nota: *Elaboración propia. La figura debe mostrar responsabilidades compartidas sin sugerir que la tecnología elimina la responsabilidad de las organizaciones.*



Capítulo 10.

Gobernanza global de la IA y perspectivas futuras del derecho



Introducción

La expansión de la inteligencia artificial plantea una cuestión que atraviesa los temas estudiados en esta obra: cómo orientar sistemas que operan a través de fronteras, intervienen en ámbitos decisivos de la vida social y cambian con rapidez, sin debilitar los derechos fundamentales ni cerrar posibilidades de innovación. La respuesta exige articular distintos niveles de gobernanza, actualizar las herramientas jurídicas existentes y construir capacidades públicas para evaluar impactos, exigir responsabilidades y ofrecer remedios. Este capítulo integra esos elementos y propone una agenda centrada en las personas, con atención a la cooperación internacional y a las condiciones institucionales de América Latina.

10.1. Gobernanza multinivel y cooperación internacional

La gobernanza global de la inteligencia artificial comprende las normas, instituciones, procedimientos y prácticas mediante los cuales los actores públicos y privados orientan el ciclo de vida de los sistemas de IA. No equivale a un gobierno mundial ni a un único código normativo. Se configura mediante la interacción entre el derecho internacional, los ordenamientos nacionales, las organizaciones regionales, los estándares técnicos y las políticas internas de empresas y entidades públicas. Esta estructura multinivel responde a una característica central de la tecnología: su desarrollo, entrenamiento, distribución y uso pueden involucrar actores y datos situados en países distintos, mientras sus efectos se producen localmente en decisiones sobre empleo, crédito, salud, educación, justicia, seguridad y acceso a servicios.

El carácter transnacional de la IA no elimina la responsabilidad de los Estados. Los compromisos en materia de derechos humanos siguen siendo relevantes cuando una autoridad utiliza sistemas automatizados o permite que su uso afecte a personas bajo su jurisdicción. La contratación de un proveedor extranjero no transfiere la obligación estatal de garantizar legalidad, igualdad, privacidad, debido proceso y acceso a recursos efectivos. Del mismo modo, las empresas conservan deberes asociados a su función y a su capacidad de prevenir o mitigar impactos adversos. La gobernanza multinivel debe, por tanto, establecer responsabilidades complementarias y evitar que la complejidad de la cadena tecnológica se convierta en una excusa para diluirlas.

En el plano internacional conviven instrumentos de distinta naturaleza jurídica. El Reglamento de Inteligencia Artificial de la Unión Europea constituye derecho vinculante para los sujetos y territorios comprendidos en su ámbito de aplicación, con obligaciones diferenciadas según el tipo de sistema, el riesgo y el papel de cada operador. Su influencia puede extenderse más allá de Europa cuando proveedores globales ajustan productos y procesos para atender ese mercado. Sin embargo, sus reglas no sustituyen automáticamente los

regímenes de protección de datos, consumo, competencia, propiedad intelectual y responsabilidad civil aplicables en cada jurisdicción. La regulación de IA se inserta en un entramado jurídico más amplio, tal como se examinó en los capítulos 4, 5 y 9.

El Consejo de Europa aprobó en 2024 la Convención Marco sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho, primer tratado internacional jurídicamente vinculante dedicado específicamente a este campo. Su propósito es que las actividades comprendidas en el ciclo de vida de la IA sean compatibles con los derechos humanos, la democracia y el Estado de derecho, a la vez que se favorece el progreso tecnológico. El instrumento tiene relevancia para Estados europeos y para los países no miembros invitados a participar en su elaboración. Su entrada en vigor y sus efectos para cada Estado dependen de las reglas de ratificación y de la situación de cada parte, por lo que no debe confundirse su adopción con una obligación universal ya vigente.

A su vez, la Recomendación de la OCDE sobre Inteligencia Artificial, revisada en 2024, proporciona principios y orientaciones de política pública para el desarrollo de una IA confiable. Incluye el respeto al Estado de derecho, los derechos humanos y los valores democráticos; la transparencia y la explicabilidad; la robustez y seguridad; y la rendición de cuentas. También insta a los gobiernos a cooperar, intercambiar conocimiento, promover estándares técnicos interoperables y construir indicadores comparables. Es un instrumento de derecho blando: orienta políticas y facilita convergencias, pero no equivale por sí mismo a un tratado ni reemplaza leyes nacionales.

La Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO, adoptada en 2021 por sus Estados miembros, refuerza una perspectiva universal que articula valores, principios y áreas de acción, incluyendo derechos humanos, inclusión, diversidad, medioambiente y gobernanza de datos. Al igual que la Recomendación de la OCDE,

no es un tratado. Su relevancia reside en ofrecer un marco común para que los Estados evalúen sus capacidades, diseñen políticas y traduzcan compromisos éticos en medidas institucionales. Su carácter voluntario no la hace irrelevante: puede influir en legislación, estrategias nacionales, contratación pública, formación y prácticas de organizaciones. Pero la referencia a principios no basta si no se acompaña de autoridades responsables, presupuestos, procedimientos de supervisión y remedios efectivos.

La cooperación también se expresa en espacios multilaterales de diálogo y producción de conocimiento. El Pacto Digital Global, incluido en el Pacto para el Futuro adoptado por la Asamblea General de las Naciones Unidas en 2024, planteó una visión compartida para la cooperación digital, con atención a la gobernanza de datos, la reducción de brechas y la participación de diversos actores. En 2025, la Asamblea General estableció mecanismos de Naciones Unidas vinculados con el Panel Científico Internacional Independiente sobre IA y el Diálogo Global sobre la Gobernanza de la IA. La primera sesión del Diálogo tuvo lugar en Ginebra en julio de 2026; reunió a delegaciones de 163 países, junto con representantes de gobiernos, sociedad civil, academia, sector privado y comunidades técnicas. La amplitud de esa participación muestra el valor de una deliberación global, aunque sus conclusiones no sustituyen instrumentos jurídicos ni garantizan por sí mismas acuerdos vinculantes.

La gobernanza multinivel enfrenta, además, una cuestión de legitimidad. Si las reglas se elaboran solo en los países con mayor capacidad tecnológica, quienes tienen menos recursos pueden quedar obligados a implementar estándares que no contribuyeron a definir. Esa asimetría es especialmente significativa para América Latina, donde los Estados pueden depender de infraestructura, modelos y servicios externos, pero deben responder por sus efectos sobre las personas. La cooperación internacional debería, por ello, incluir transferencia de conocimiento, formación técnica, acceso a capacidad computacional y participación

efectiva en órganos de normalización. También debe facilitar que los países compartan evaluaciones e incidentes sin obligarlos a adoptar mecánicamente un único modelo regulatorio.

La interoperabilidad jurídica es una meta útil si se entiende como compatibilidad de salvaguardas y resultados, no como uniformidad completa. Los instrumentos internacionales pueden acordar principios comunes y vocabularios mínimos, mientras las legislaciones nacionales ajustan procedimientos y obligaciones a sus estructuras constitucionales. Para que esa compatibilidad sea concreta, es necesario trabajar en definiciones, clasificación de riesgos, documentación, reporte de incidentes, técnicas de auditoría, protección de datos y cooperación entre autoridades. Los estándares técnicos pueden apoyar esta labor, pero deben someterse a deliberación pública y control jurídico. Una norma de consenso técnico no puede decidir por sí sola el alcance de un derecho ni cerrar vías de impugnación.

La arquitectura institucional debe asignar funciones sin producir vacíos ni duplicaciones. Las autoridades de protección de datos, consumo, competencia, telecomunicaciones, trabajo, salud y supervisión financiera ya poseen competencias relevantes para evaluar usos de IA en sus sectores. Puede ser necesario crear organismos especializados o unidades de coordinación, pero una autoridad central no necesariamente reemplaza a reguladores sectoriales con conocimiento del contexto. La opción más sólida suele combinar coordinación transversal, especialización sectorial, intercambio de información y reglas claras de remisión de casos. También requiere independencia, personal técnico y jurídico, acceso a evidencia y capacidad de emitir decisiones motivadas.

Así entendida, la cooperación internacional no persigue únicamente armonizar textos normativos. Debe facilitar la protección efectiva de las personas, en especial cuando una decisión automatizada atraviesa fronteras o cuando los responsables se encuentran fuera del

país afectado. Mecanismos de asistencia, notificación y cooperación regulatoria pueden reducir la distancia entre el daño y la autoridad capaz de investigarlo. La persona afectada debe poder identificar quién desplegó el sistema, conocer el canal de reclamación y obtener una revisión con efectos prácticos. El nivel global proporciona principios y cooperación; el regional, reglas comunes; el nacional, autoridades y garantías; y el institucional, controles concretos sobre cada sistema.

Tabla 10.1

Instrumentos de gobernanza de IA y su naturaleza

Instrumento o espacio	Nivel	Naturaleza jurídica	Función principal	Cautela para su aplicación
Reglamento de Inteligencia Artificial de la Unión Europea	Regional	Norma vinculante dentro de su ámbito de aplicación	Establece obligaciones según el tipo de sistema, el riesgo y la función de cada actor	No es una norma universal y debe aplicarse junto con otras leyes pertinentes
Convención Marco del Consejo de Europa sobre IA y derechos humanos	Internacional	Tratado sujeto a firma, ratificación y entrada en vigor para cada Estado	Vincula las actividades relacionadas con la IA con los derechos humanos, la democracia y el Estado de derecho	Su adopción no significa que ya sea obligatoria para todos los países
Recomendación de la OCDE sobre IA	Internacional	Recomendación intergubernamental	Orienta políticas para promover una IA confiable, respetuosa de los derechos y responsable	No equivale por sí sola a una ley nacional ni a un tratado

Instrumento o espacio	Nivel	Naturaleza jurídica	Función principal	Cautela para su aplicación
Recomendación de la UNESCO sobre la Ética de la IA	Internacional	Recomendación	Ofrece principios y áreas de acción para políticas centradas en las personas, los derechos humanos y la sostenibilidad	Requiere medidas nacionales e institucionales para llevar sus principios a la práctica
Pacto Digital Global de las Naciones Unidas	Global	Compromiso político internacional	Promueve cooperación digital, reducción de brechas y una visión compartida sobre gobernanza tecnológica	No sustituye las leyes nacionales ni crea por sí solo un régimen vinculante específico de IA
Diálogo Global de las Naciones Unidas sobre Gobernanza de la IA	Global	Espacio de diálogo y cooperación	Facilita el intercambio entre Estados y otros actores sobre oportunidades, riesgos y políticas de IA	Sus debates y conclusiones no equivalen por sí mismos a normas jurídicas obligatorias

Nota. Elaboración propia con base en los instrumentos y espacios de cooperación examinados en el capítulo. La fuerza jurídica y los efectos de cada instrumento dependen de su naturaleza, su ámbito de aplicación y, cuando corresponda, de la ratificación e implementación nacional.

10.2. IA generativa, sistemas autónomos y nuevos desafíos regulatorios

La IA generativa ha modificado las condiciones de producción y

circulación de textos, imágenes, audio, video y código. Los modelos de propósito general pueden ser incorporados en múltiples servicios y tareas, muchas veces sin que el usuario conozca el proveedor original, los datos de entrenamiento, las limitaciones del sistema o los cambios introducidos en una nueva versión. Esta capacidad de reutilización amplía la innovación, pero complica la evaluación jurídica: un mismo modelo puede tener usos de bajo impacto en una aplicación y consecuencias graves en otra. Por ello, la regulación debe considerar tanto las características del modelo como el propósito, el contexto de despliegue y la exposición concreta a riesgos.

Uno de los desafíos es la fiabilidad. Los modelos generativos producen resultados plausibles, pero pueden incluir información incorrecta o inventada, omitir contexto, reproducir sesgos o presentar inferencias como hechos. En ámbitos jurídicos, sanitarios, educativos o administrativos, el uso de tales resultados sin verificación puede afectar derechos y decisiones importantes. La advertencia general de que una salida puede ser errónea no siempre basta; la organización debe definir qué tareas requieren revisión humana cualificada, qué fuentes deben corroborarse y qué decisiones no se delegan. La supervisión humana debe tener autoridad y condiciones materiales para corregir, rechazar o detener el sistema, no ser una formalidad que legitime automáticamente una recomendación algorítmica.

Otro problema se refiere a la autenticidad y la integridad informativa. La creación de imágenes, voces y videos sintéticos puede facilitar la expresión artística y la accesibilidad, pero también permite fraude, suplantación, manipulación de evidencia y campañas de desinformación. Las técnicas de etiquetado, procedencia y marca de agua pueden ayudar a informar sobre el origen de contenidos, aunque no son infalibles ni deben convertirse en una exigencia general de vigilancia sobre toda expresión. Las medidas deben atender al contexto y al daño: una interacción cotidiana con un asistente, una obra artística y un material presentado como prueba requieren análisis distintos.

Las restricciones deben ser legales, necesarias y proporcionales, con garantías frente a la censura y mecanismos de apelación.

La protección de datos y la propiedad intelectual siguen siendo cuestiones centrales. El entrenamiento puede involucrar información personal, obras protegidas y otros contenidos cuya recopilación, transformación o uso generen controversias jurídicas. La relación entre minería de textos y datos, excepciones al derecho de autor, licencias, transparencia de datos de entrenamiento y posibilidad de reservar obras varía entre jurisdicciones. No existe una regla universal que resuelva de modo automático todos los casos. Como se estudió en el capítulo 9, el análisis debe distinguir la titularidad de la salida, los derechos sobre los insumos, las condiciones de uso del modelo y la responsabilidad de quienes despliegan el sistema. En el capítulo 4 se examinó, además, que los datos personales mantienen su protección aunque se incorporen a procesos de entrenamiento o generación.

La clasificación de riesgos debe adaptarse a sistemas de propósito general sin borrar la diversidad de situaciones. Una evaluación concentrada únicamente en el desarrollador del modelo puede ignorar los riesgos introducidos por integradores, proveedores de aplicaciones o entidades que deciden desplegarlo. En sentido inverso, exigir al usuario final que identifique y gestione riesgos que no puede observar ni controlar tampoco es razonable. La distribución de obligaciones debe considerar conocimiento, acceso a información, capacidad de modificación, control sobre la finalidad y posibilidad de prevenir daños. Para modelos generales, resultan relevantes la documentación técnica, pruebas de capacidades y limitaciones, políticas de uso, información sobre riesgos sistémicos y canales de notificación. Para el despliegue, deben valorarse finalidad, población afectada, datos utilizados, medidas de supervisión y efectos previsibles.

La evolución hacia sistemas más autónomos intensifica estas cuestiones. Un agente de IA puede recibir un objetivo, planificar pasos, consultar

herramientas, ejecutar código o interactuar con servicios externos con una intervención humana reducida. La autonomía puede ser útil en tareas administrativas y de análisis, pero también amplía la superficie de error: el sistema puede actuar con información incompleta, interpretar mal una instrucción, acceder a recursos no previstos o encadenar decisiones que generen efectos acumulativos. La regulación debe atender a la capacidad real de actuar y a los mecanismos de autorización, registro, limitación y reversibilidad. La etiqueta "autónomo" no crea personalidad jurídica ni elimina la responsabilidad de las personas y organizaciones que diseñan, integran o autorizan el sistema.

Los sistemas que interactúan con el mundo físico plantean riesgos adicionales. Robots, vehículos, dispositivos médicos e infraestructura automatizada pueden causar daños materiales o corporales, además de afectar seguridad y privacidad. En esos escenarios, la evaluación debe incluir pruebas bajo condiciones previsibles de uso y mal uso, protocolos de emergencia, límites de operación y mecanismos seguros de detención. La interacción con sistemas digitales de pagos, archivos, servicios públicos o comunicaciones también puede producir consecuencias significativas, aunque no exista un componente físico. La autonomía debe entenderse en relación con el grado de intervención, la irreversibilidad de las acciones, la gravedad potencial del daño y la posibilidad de una respuesta humana oportuna.

Los riesgos de seguridad requieren una perspectiva de ciclo de vida. La extracción de datos, la manipulación de entradas, la inyección de instrucciones, la filtración de información confidencial y el uso de capacidades para actividades dañinas pueden aparecer después del despliegue, incluso si el sistema superó pruebas iniciales. La robustez no es una propiedad permanente certificada una sola vez. Depende de actualizaciones, integración con otras herramientas, cambios en los datos y condiciones de uso. Las organizaciones deberían conservar registros proporcionales al riesgo, mantener procesos para reportar incidentes y revisar medidas cuando se producen fallas relevantes. La

divulgación responsable es necesaria, pero no debe publicar detalles que faciliten la explotación de vulnerabilidades.

El impacto laboral también merece atención. La IA generativa puede automatizar tareas de redacción, búsqueda, traducción, atención al cliente y análisis, al tiempo que crea nuevas funciones y modifica el contenido de las ocupaciones existentes. El efecto no se reduce al reemplazo de puestos; comprende intensificación del trabajo, vigilancia, control de productividad, cambios en autonomía profesional y nuevas desigualdades entre quienes pueden complementar su labor con la tecnología y quienes asumen sus costos. Los sistemas utilizados para contratación, evaluación o asignación de tareas afectan derechos laborales y pueden reforzar sesgos, asunto relacionado con el análisis del capítulo 3. La respuesta pública requiere diálogo social, formación continua, protección frente a decisiones arbitrarias y distribución justa de los beneficios de productividad.

La concentración de recursos y capacidades es un desafío regulatorio transversal. El desarrollo de modelos de frontera exige datos, cómputo, talento, infraestructura y acceso a mercados que se encuentran distribuidos de manera desigual. Esa concentración puede limitar la competencia y la diversidad lingüística y cultural, aumentar la dependencia de proveedores externos y reducir la capacidad de los Estados para auditar sistemas propietarios. Las obligaciones deben ser proporcionales al rol y al riesgo, pero la proporcionalidad no debería transformarse en una exención general para actores con gran influencia sistémica. La apertura de modelos y herramientas puede favorecer investigación y adaptación local, aunque también puede facilitar usos dañinos; su evaluación debe basarse en capacidades, contexto y medidas de mitigación, no en etiquetas absolutas.

El derecho tendrá que responder a estas transformaciones sin adoptar dos extremos: considerar que toda innovación es intrínsecamente beneficiosa o presumir que todo sistema nuevo requiere una prohibición

general. La regulación debe ser capaz de intervenir antes del daño cuando existen riesgos graves y, a la vez, permitir experimentación responsable cuando las incertidumbres pueden reducirse mediante controles. La revisión de reglas debe basarse en evidencia, consulta y evaluación de efectos. Cuando cambian las capacidades o se descubren riesgos que no se anticiparon, las obligaciones pueden requerir actualización. Esta apertura adaptativa no implica inseguridad jurídica si se construye mediante criterios públicos, plazos razonables, decisiones motivadas y revisión judicial.

10.3. Hacia un marco jurídico adaptativo: sandboxes, auditorías y evaluaciones de impacto

Un marco adaptativo combina normas estables con procedimientos que permiten aprender y corregir. La rapidez del cambio tecnológico dificulta que cada obligación quede descrita en detalle legislativo, pero no justifica entregar discreción ilimitada a autoridades o empresas. El punto de equilibrio consiste en que la ley establezca derechos, responsabilidades, prohibiciones, criterios de riesgo y mecanismos de control; las normas secundarias y los estándares técnicos pueden precisar métodos de cumplimiento, siempre sujetos a transparencia, participación y revisión. Las cláusulas de evaluación periódica deben verificar si las reglas alcanzan sus objetivos y si producen efectos desproporcionados o vacíos de protección.

Los espacios controlados de experimentación, conocidos como sandboxes regulatorios, permiten que proveedores y autoridades observen el comportamiento de sistemas innovadores en condiciones delimitadas. En la Unión Europea, el Reglamento de IA prevé sandboxes supervisados por autoridades competentes para facilitar desarrollo, entrenamiento, prueba y validación durante un periodo limitado, con un plan acordado y salvaguardas. La supervisión puede identificar riesgos para derechos fundamentales, salud y seguridad, además de apoyar el cumplimiento normativo. El Reglamento también regula pruebas en condiciones reales bajo requisitos específicos. Un sandbox no equivale

a una licencia general ni suspende las obligaciones jurídicas aplicables; su legitimidad depende de límites claros, supervisión efectiva y protección de quienes puedan verse afectados.

Para evitar que el sandbox se convierta en una zona de excepción, la autoridad debe definir de antemano su objetivo, alcance, duración, participantes, población expuesta, condiciones de salida y criterios de interrupción. La participación de la entidad supervisada no elimina la independencia del regulador. Las personas potencialmente afectadas deben recibir información adecuada y contar con canales de consulta y reclamación. Si las pruebas involucran datos personales o servicios esenciales, deberán incorporarse medidas específicas de minimización, seguridad, supervisión humana y continuidad del servicio. Los resultados agregados pueden publicarse para que el aprendizaje no quede exclusivamente en manos del proveedor. Cuando una prueba produzca un daño, la existencia del sandbox no debería excluir la investigación ni el derecho a reparación.

La evaluación de impacto constituye otro instrumento preventivo. No es un formulario que se completa al final del diseño, sino un proceso para identificar quiénes pueden verse afectados, cuáles derechos están en juego, qué daños son plausibles y qué alternativas menos riesgosas existen. Debe examinar calidad y representatividad de datos, posibles sesgos, explicabilidad, seguridad, accesibilidad, contexto de uso, escalabilidad y mecanismos de impugnación. En la Unión Europea, el Reglamento de IA incorpora una evaluación de impacto sobre los derechos fundamentales para determinados responsables del despliegue de sistemas de alto riesgo, con alcance definido por la norma. Esa obligación específica debe distinguirse de las evaluaciones de impacto de protección de datos previstas en otros instrumentos, aunque ambas puedan coordinarse cuando corresponda.

La evaluación debe ser proporcional a la gravedad del uso, pero no puede quedar reducida a una autoevaluación opaca del proveedor. En

sectores como justicia, seguridad, migración, empleo, crédito y salud, la autoridad pública debería poder solicitar información suficiente para verificar los supuestos y las medidas. Las personas afectadas y grupos con conocimiento del contexto pueden identificar riesgos que no aparecen en pruebas técnicas, como barreras lingüísticas, efectos diferenciados sobre comunidades o incompatibilidad con procedimientos locales. La consulta no sustituye la decisión jurídica, pero mejora su base empírica. Un informe útil expone limitaciones, riesgos residuales, medidas de mitigación, responsables de implementarlas y condiciones para detener o modificar el sistema.

Las auditorías algorítmicas pueden aportar evidencia sobre el funcionamiento y la gobernanza del sistema. Una auditoría técnica examina elementos como datos, rendimiento, robustez y seguridad. Una auditoría de derechos evalúa impactos en igualdad, privacidad, libertad, acceso a servicios y garantías procesales. Una auditoría organizacional revisa responsabilidades internas, contratación, documentación, capacitación y respuesta a incidentes. Estas perspectivas son complementarias. Ningún indicador estadístico, por sí solo, determina si un sistema es jurídicamente aceptable; los resultados deben interpretarse respecto de la finalidad, el contexto y la distribución de beneficios y daños. La auditoría tampoco debe limitarse a comprobar que existen políticas escritas: debe verificar si los controles se aplican y si permiten corregir efectos adversos.

La independencia y la competencia de quienes auditan son condiciones esenciales. Si la entidad auditada selecciona al evaluador, limita el acceso a datos o financia el trabajo sin reglas de transparencia, puede surgir un conflicto de interés. Los procedimientos deben especificar acceso seguro a documentación, registros, resultados de prueba y personal pertinente, con salvaguardas para secretos comerciales y datos personales. La confidencialidad no debe impedir que la autoridad conozca la evidencia necesaria ni que el público reciba información suficiente sobre riesgos y decisiones. Los informes pueden ofrecer

niveles de detalle: una versión técnica para supervisores, una versión pública comprensible y, cuando sea necesario, anexos protegidos. El secreto empresarial no es una respuesta automática frente a solicitudes legales de información.

La supervisión posterior al despliegue complementa la evaluación previa. Un sistema puede funcionar de manera distinta cuando se integra con otras herramientas o se expone a poblaciones y condiciones no presentes en la prueba inicial. Las organizaciones deben establecer indicadores de desempeño y daño, mecanismos de recepción de quejas, análisis de fallas y revisión de actualizaciones. El monitoreo debe detectar cambios de distribución, tasas desiguales de error, degradación del rendimiento y usos que exceden la finalidad declarada. La obligación de monitorear no autoriza la recopilación ilimitada de datos ni la vigilancia indiscriminada de usuarios. Se requiere una base jurídica, finalidades definidas y medidas de minimización y seguridad.

Los mecanismos de notificación de incidentes ayudan a que la información relevante llegue oportunamente a autoridades y usuarios. Para ser eficaces, deben definir qué eventos se notifican, a quién, en qué plazo y con qué contenido, según la gravedad y el tipo de sistema. Una falla crítica con riesgo de daño físico no debe tratarse igual que un error menor sin impacto material. Los reportes pueden incluir contexto, número y tipo de personas afectadas, medidas de contención, causas preliminares y acciones correctivas. La información compartida entre organismos nacionales e internacionales puede revelar patrones que un proveedor o una autoridad aislada no detectaría. La confidencialidad puede proteger datos sensibles, pero no debe invisibilizar incidentes sistémicos.

El régimen adaptativo también debe conectarse con contratación pública y responsabilidad. Antes de adquirir un sistema, una entidad estatal puede exigir documentación, acceso a pruebas, trazabilidad, actualización, interoperabilidad y asistencia para atender

reclamaciones. Las condiciones contractuales deberían indicar quién responde ante errores, qué información debe conservarse, cómo se notifican cambios sustanciales y en qué casos puede suspenderse el servicio. La administración no puede delegar la responsabilidad por una decisión pública en cláusulas que impidan explicar o revisar el resultado. Este enfoque enlaza con el capítulo 8 sobre usos estatales de IA y con el capítulo 7 sobre atribución y reparación por daños.

La evidencia obtenida en evaluaciones, auditorías y sandboxes debería alimentar reglas y decisiones futuras. Para ello, es útil mantener registros de sistemas desplegados en sectores de mayor impacto, publicar estadísticas agregadas y documentar decisiones de autorización, restricción o retirada. Los registros deben equilibrar transparencia con seguridad, privacidad y secretos legítimos. También conviene que las autoridades expliquen qué criterios aplicaron y cómo resolvieron discrepancias técnicas. Esa práctica genera previsibilidad y permite a jueces, investigadores y sociedad civil examinar la consistencia de la regulación.

La adaptabilidad jurídica exige conservar vías de impugnación. Una autoridad debe poder exigir correcciones, suspender un uso o prohibirlo cuando el riesgo no pueda mitigarse. Una persona afectada debe poder solicitar revisión humana, conocer fundamentos relevantes y acceder a un procedimiento imparcial. La posibilidad de corregir el sistema no reemplaza la reparación de un daño ya ocurrido. En su conjunto, los instrumentos adaptativos son legítimos cuando fortalecen el control democrático y la tutela de derechos; pierden legitimidad cuando se usan para postergar indefinidamente decisiones, normalizar la experimentación sobre poblaciones vulnerables o trasladar los riesgos al público sin mecanismos de compensación.

Figura 10.1.

Marco adaptativo de gobernanza durante el ciclo de vida de la IA



Nota: *Elaboración propia con base en el Reglamento (UE) 2024/1689 y las recomendaciones de la OCDE y la UNESCO.*

10.4. Recomendaciones y agenda de política pública para una IA centrada en las personas

Una agenda pública centrada en las personas debe comenzar con una premisa jurídica: la IA es un medio sujeto a fines legítimos y a límites constitucionales, no una autoridad autónoma que determina por sí misma qué es justo o conveniente. Las decisiones sobre derechos, distribución de recursos y ejercicio del poder requieren responsables identificables, razones comprensibles y vías de revisión. Este principio no impide automatizar tareas rutinarias ni aprovechar capacidades analíticas. Exige que la intensidad del control aumente cuando el sistema afecta oportunidades esenciales, personas en situación de vulnerabilidad o libertades públicas.

La primera prioridad consiste en consolidar un piso jurídico de garantías. Cada Estado debería identificar qué normas existentes ya se aplican a los sistemas de IA en protección de datos, igualdad, consumo,

trabajo, propiedad intelectual, contratación y responsabilidad. Donde existan vacíos demostrables, la legislación debe establecer obligaciones claras para quienes desarrollan, proveen y despliegan sistemas, atendiendo a sus funciones y capacidades. Las prohibiciones deberían reservarse para prácticas incompatibles con derechos o que generen riesgos inadmisibles, y definirse con suficiente precisión para prevenir arbitrariedad. Las obligaciones de transparencia, supervisión, documentación y gestión de riesgos deben ser proporcionales y verificables.

La segunda prioridad es construir capacidad institucional. Los reguladores necesitan especialistas jurídicos, técnicos, estadísticos y sociales, además de recursos para acceder a peritajes independientes y mantener herramientas propias de evaluación. La inversión no debe concentrarse en una oficina central sin coordinación sectorial. Se recomienda establecer mecanismos permanentes entre autoridades y protocolos de consulta para casos que involucren más de una competencia. Las instituciones públicas también deben formar a jueces, defensorías, fiscalías, órganos de contratación y funcionarios que utilizan sistemas automatizados. Sin conocimiento técnico mínimo, resulta difícil valorar una explicación, detectar sesgos o determinar si la intervención humana fue significativa.

En América Latina, la agenda debe reconocer diferencias de capacidad estatal, infraestructura, diversidad lingüística y disponibilidad de datos representativos. No resulta suficiente importar modelos entrenados en otros contextos y asumir que sus métricas se trasladan sin cambios. Los sistemas pueden clasificar de manera distinta cuando el lenguaje, las instituciones, los patrones sociales o las condiciones de acceso difieren. Antes de desplegar una herramienta, las autoridades deberían verificar su pertinencia local, evaluar la calidad de los datos y consultar a comunidades potencialmente afectadas. Las estrategias nacionales deben conectarse con cooperación regional para compartir conocimiento, laboratorios de prueba, formación y criterios de

evaluación.

La tercera prioridad es orientar el poder de compra público. Los Estados son compradores relevantes de tecnología y pueden exigir condiciones que mejoren la calidad de mercado. La contratación debería especificar la finalidad permitida, las métricas de desempeño, las obligaciones de actualización, la documentación disponible, las salvaguardas de datos, el acceso de auditores y las responsabilidades ante fallas. Los pliegos no deberían pedir únicamente que un proveedor declare que su producto es "ético" o "confiable". Deben establecer pruebas y evidencia verificables. En sistemas de alto impacto, puede ser necesario realizar una evaluación antes de la compra y revisarla durante la operación, además de publicar información sobre el uso y las vías de reclamación.

La cuarta prioridad es hacer efectiva la participación de las personas. Las consultas públicas deben incorporar no solo a empresas y especialistas, sino también a sindicatos, asociaciones profesionales, organizaciones de derechos humanos, comunidades indígenas, personas con discapacidad, consumidores y grupos afectados por decisiones automatizadas. La participación debe ocurrir en etapas tempranas, cuando todavía es posible cambiar la finalidad, elegir otra tecnología o decidir que no se utilizará IA. También debe mantenerse durante la evaluación de impactos y la revisión de políticas. No toda objeción puede resolverse de manera consensuada, pero la autoridad tiene que explicar cómo ponderó los aportes y qué medidas adoptó frente a riesgos señalados.

La quinta prioridad es garantizar acceso a explicación, revisión y reparación. Las personas deben poder saber cuándo una decisión relevante se apoyó en IA, cuál entidad es responsable y cómo solicitar revisión. Una explicación adecuada no necesariamente requiere revelar cada parámetro del modelo; debe proporcionar información que permita comprender los elementos decisivos y controvertir errores o discriminación. La revisión humana debe ser realizada por una

persona con competencia, tiempo y autoridad suficientes para cambiar el resultado. Si hubo daño, las vías de reparación deben ser accesibles, oportunas y capaces de considerar la complejidad de la cadena tecnológica. La opacidad no debería obligar a la persona a demostrar hechos a los que solo el proveedor tiene acceso.

La sexta prioridad es promover competencia, investigación y diversidad. Una gobernanza centrada en derechos no puede desatender la concentración económica ni la dependencia tecnológica. Las políticas deberían apoyar investigación pública, repositorios y herramientas abiertas con salvaguardas, capacidad de cómputo accesible y colaboración entre universidades, sector público y actores locales. Los estándares abiertos e interoperables pueden reducir dependencia de proveedores y facilitar auditoría. Sin embargo, la apertura no constituye automáticamente una garantía de seguridad ni de equidad. Las decisiones sobre modelos abiertos o cerrados deben considerar contexto, capacidades de uso, riesgos de abuso y mecanismos para mitigar daños. La competencia y la protección de derechos son objetivos compatibles cuando se diseñan obligaciones proporcionales y previsibles.

La séptima prioridad es desarrollar una política de datos legítima y sostenible. La disponibilidad de datos de calidad puede ayudar a construir sistemas más adecuados, pero la recolección masiva no es un objetivo en sí mismo. Las políticas deben respetar finalidad, minimización, seguridad, derechos de las personas y gobernanza de comunidades. Los proyectos que usan información sensible requieren justificación reforzada y medidas para evitar reidentificación o usos secundarios incompatibles. También deben considerar los costos ambientales de centros de datos y modelos de gran escala. La sostenibilidad comprende energía, agua, extracción de materiales, gestión de equipos y efectos sociales de las infraestructuras. Las decisiones públicas pueden promover medición transparente y criterios de eficiencia sin convertir una métrica aislada en sustituto del análisis

de derechos.

La octava prioridad es fortalecer alfabetización y capacidades sociales. La alfabetización en IA no consiste únicamente en enseñar a utilizar herramientas. Debe permitir comprender límites, sesgos, incertidumbre, privacidad, seguridad y mecanismos de impugnación. También es necesario preparar a profesionales para trabajar con sistemas de IA sin renunciar al juicio especializado. En el ámbito jurídico, esto comprende entender documentación técnica, formular preguntas pertinentes a peritos, valorar evidencia automatizada y explicar a las personas las consecuencias de una decisión. La actualización profesional debe abarcar operadores judiciales, abogados, funcionarios, desarrolladores, directivos y trabajadores. Los programas han de evitar que la responsabilidad de adaptarse recaiga exclusivamente en quienes enfrentan cambios laborales.

La novena prioridad es institucionalizar evaluación y revisión. Cada política nacional debería establecer objetivos verificables: reducción de daños, acceso efectivo a recursos, mejora de servicios, disminución de brechas y distribución justa de beneficios. La evaluación debe combinar indicadores cuantitativos con análisis cualitativo y revisión de casos. Un sistema puede cumplir una métrica técnica y, aun así, producir resultados injustos por su contexto de uso. Los indicadores deben desagregarse cuando sea legal y éticamente apropiado, para identificar impactos diferenciados sin crear nuevas exposiciones de datos. Las normas deberían revisarse periódicamente, pero los cambios necesitan consulta, análisis de impacto regulatorio y periodos de adaptación para preservar seguridad jurídica.

La décima prioridad es asumir la cooperación internacional como una política de capacidad compartida. Los países latinoamericanos pueden participar en foros globales y regionales con posiciones coordinadas, promover estándares compatibles con los derechos humanos y solicitar apoyo para investigación, formación y supervisión. La cooperación

debería incluir intercambio de incidentes y buenas prácticas, asistencia técnica a autoridades, acceso a recursos de evaluación y apoyo para sistemas en lenguas y contextos insuficientemente representados. Los acuerdos sobre IA deben evitar que la interoperabilidad se convierta en una vía para rebajar garantías o condicionar el acceso a servicios esenciales a tecnologías no evaluadas localmente.

El diseño de escenarios futuros puede organizarse en torno a tres trayectorias. En una primera, predomina la fragmentación: los Estados adoptan reglas incompatibles, las empresas eligen jurisdicciones más favorables y las personas enfrentan niveles distintos de protección. En una segunda, se produce una armonización mínima que mejora el intercambio de información y fija obligaciones comunes, aunque deja brechas de implementación. En una tercera, la cooperación combina estándares globales de derechos, marcos regionales compatibles y capacidad nacional para supervisar y reparar. Esta última trayectoria no surge automáticamente de la tecnología ni de las declaraciones de principios; requiere decisiones presupuestarias, instituciones independientes y participación social sostenida.

La transición hacia ese escenario demanda una secuencia práctica. Primero, inventariar los sistemas utilizados por el Estado y los sectores regulados, con atención prioritaria a usos de alto impacto. Segundo, asignar una autoridad responsable y definir estándares de evaluación, documentación e incidentes. Tercero, revisar contratos y prácticas de adquisición para garantizar acceso a información y posibilidad de control. Cuarto, establecer canales de reclamación, revisión y reparación. Quinto, evaluar resultados y publicar aprendizajes de manera que permitan ajustar la política. Esta secuencia puede adaptarse a las condiciones de cada país, pero debe conservar el núcleo: la responsabilidad permanece en las instituciones y actores humanos, incluso cuando parte del proceso se automatiza.

La IA centrada en las personas no se mide por la cantidad de sistemas

desplegados ni por su sofisticación técnica, sino por la legitimidad de sus fines, la distribución de sus beneficios, la prevención de daños y la capacidad de las personas para ejercer sus derechos. El Derecho tiene una función de garantía y también de orientación: crea condiciones para que la innovación sirva al bienestar común sin normalizar decisiones opacas o poderes incontestables. La gobernanza global será legítima si hace posible que los Estados cooperen sin sacrificar su responsabilidad, que las empresas innoven bajo obligaciones verificables y que quienes resultan afectados puedan conocer, cuestionar y corregir decisiones. En esa tarea converge el argumento de la obra: el desarrollo tecnológico y los derechos fundamentales no son objetivos incompatibles, pero su compatibilidad requiere instituciones, reglas y controles efectivos.

BIBLIOGRAFÍA

Comisión Interamericana de Derechos Humanos. (2000). *Declaración de principios sobre libertad de expresión*. Organización de los Estados Americanos. https://www.oas.org/es/cidh/expresion/showarticle.asp?artID=26&lID=2

Comisión Interamericana de Derechos Humanos. (s. f.). *Principios actualizados sobre la privacidad y la protección de datos personales*. Organización de los Estados Americanos. https://www.oas.org/es/sla/ddi/proteccion_datos_personales.asp

Comunidad Andina. (1993). *Decisión 351: Régimen común sobre derecho de autor y derechos conexos*. https://www.comunidadandina.org/StaticFiles/DocOf/DEC351.pdf

Consejo de Europa. (2024). *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* (CETS No. 225). https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence

Corte Interamericana de Derechos Humanos. (1969). *Convención Americana sobre Derechos Humanos*. Organización de los Estados Americanos. https://www.oas.org/dil/esp/1969_Convenci%C3%B3n_Americana_sobre_Derechos_Humanos.pdf

European Parliament and Council of the European Union. (2016). *Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Official Journal of the European Union, L 119*, 1–88. https://eur-lex.europa.eu/eli/reg/2016/679/oj

European Parliament and Council of the European Union. (2019). *Directive (EU) 2019/790 of 17 April 2019 on copyright and related rights in the Digital Single Market. Official Journal of the European Union, L 130*, 92–125. https://eur-lex.europa.eu/eli/dir/2019/790/oj

European Parliament and Council of the European Union. (2022). *Regulation (EU) 2022/1925 of 14 September 2022 on contestable and fair markets in the digital sector (Digital Markets Act). Official Journal of the European Union, L 265*, 1–66. https://eur-lex.europa.eu/eli/reg/2022/1925/oj

European Parliament and Council of the European Union. (2024a). *Directive (EU) 2024/2853 of 23 October 2024 on liability for defective products. Official Journal of the European Union, L 2024/2853*. https://eur-lex.europa.eu/eli/dir/2024/2853/oj

European Parliament and Council of the European Union. (2024b). *Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, L 2024/1689*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Federal Trade Commission. (2025). *Partnerships between cloud service providers and AI developers: FTC staff report*. https://www.ftc.gov/news-events/news/press-releases/2025/01/ftc-issues-staff-report-ai-partnerships-investments-study

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). https://doi.org/10.1162/99608f92.8cd550d1

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*, 689–707. https://doi.org/10.1007/s11023-018-9482-5

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg,

J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review, 165*(3), 633–705.

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology, 6*, 175–183. https://doi.org/10.1007/s10676-004-3422-1

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2). https://doi.org/10.1177/2053951716679679

Naciones Unidas. (1966). *Pacto Internacional de Derechos Civiles y Políticos*. https://www.ohchr.org/es/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights

Naciones Unidas. (2006). *Convención sobre los derechos de las personas con discapacidad*. https://www.ohchr.org/es/instruments-mechanisms/instruments/convention-rights-persons-disabilities

Naciones Unidas. (2024). *Pacto Digital Global* (Anexo I de la Resolución A/RES/79/1). https://www.un.org/global-digital-compact/

Naciones Unidas. (2025). *Establishment of an independent international scientific panel on artificial intelligence and a global dialogue on artificial intelligence governance* (A/RES/79/325). https://docs.un.org/en/A/RES/79/325

Naciones Unidas. (2026). *Co-Chairs' summary from the inaugural Global Dialogue on AI Governance: Geneva, 6–7 July 2026*. https://www.un.org/global-dialogue-ai-governance/sites/default/files/2026-09/ai_dialogue_co-chairs_summary.pdf

National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1

Organisation for Economic Co-operation and Development.

(2024a). *Explanatory memorandum on the updated OECD definition of an AI system*. OECD Artificial Intelligence Papers. https://doi.org/10.1787/623da898-en

Organisation for Economic Co-operation and Development. (2024b). *Revised recommendation of the Council on Artificial Intelligence* (C/MIN(2024)16/FINAL). https://one.oecd.org/document/C/MIN(2024)16/FINAL/en

República del Ecuador. (2008). *Constitución de la República del Ecuador*. Registro Oficial 449. https://www.asambleanacional.gob.ec/sites/default/files/documents/old/constitucion_de_bolsillo.pdf

República del Ecuador. (2021). *Ley Orgánica de Protección de Datos Personales*. Registro Oficial, Quinto Suplemento No. 459. https://www.asambleanacional.gob.ec/sites/default/files/private/asambleanacional/filesasambleanacionalnameuid-29/Leyes%202013-2017/920-lmoreno/datos_personales.pdf

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. En *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 33–44). Association for Computing Machinery. https://doi.org/10.1145/3351095.3372873

Suprema Corte de Wisconsin. (2016). *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016). https://law.justia.com/cases/wisconsin/supreme-court/2016/2015ap000157-cr.html

United Nations Commission on International Trade Law. (1996). *UNCITRAL Model Law on Electronic Commerce*. https://uncitral.un.org/en/texts/ecommerce/modellaw/electronic_commerce

United Nations Commission on International Trade Law. (2024). *Ley Modelo de la CNUDMI sobre Contratación Automatizada*. https://uncitral.un.org/es/mlac

UNESCO. (2021). *Recomendación sobre la ética de la inteligencia artificial*. https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa

United States Copyright Office. (2023). *Copyright registration guidance: Works containing material generated by artificial intelligence*. https://www.copyright.gov/ai/ai_policy_guidance.pdf

United States Copyright Office. (2025a). *Copyright and artificial intelligence, Part 2: Copyrightability*. https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf

United States Copyright Office. (2025b). *Copyright and artificial intelligence, Part 3: Generative AI training* [Versión de prepublicación]. https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf

World Intellectual Property Organization. (s. f.). *Artificial intelligence and intellectual property*. https://www.wipo.int/en/web/frontier-technologies/artificial-intelligence/index

Fuentes del marco peruano citadas en el capítulo 6

República del Perú. (2023). *Ley n.º 31814, Ley que promueve el uso de la inteligencia artificial en favor del desarrollo económico y social del país*. https://busquedas.elperuano.pe/dispositivo/NL/2192926-1

República del Perú. (2025). *Decreto Supremo n.º 115-2025-PCM, que aprueba el Reglamento de la Ley n.º 31814*. *Diario Oficial El Peruano*. https://busquedas.elperuano.pe/



MGTR. LUIS ANTONIO GUERRA RODRÍGUEZ

c27917@utp.edu.pe

https://orcid.org/0000-0002-5472-7163

Abogado por la Universidad Privada Antenor Orrego de Trujillo – Perú, Maestro en Gestión Pública por la Universidad César Vallejo de Trujillo – Perú, Maestrando en Derecho Penal y Procesal Penal por la Universidad César Vallejo de Trujillo – Perú, Docente Universitario en la Escuela de Pregrado en la Universidad Tecnológica del Perú – Filial Chimbote y en la Universidad Nacional del Santa en la Escuela de Postgrado, Asistente Judicial en el Módulo Corporativo de Familia de la Corte Superior de Justicia del Santa.



BRAYHAN MARTÍN LEYTON COTRINA

bcotrinaleyton@gmail.com
https://orcid.org/0009-0004-3797-9491

Abogado por la Universidad Privada Antenor Orrego de Trujillo – Perú, Maestrando en Derecho Penal por la Universidad San Martín de Porres de Lima - Perú, especialista en Derecho Penal y Procesal Penal, con experiencia de cinco años en litigios en materia penal, socio fundador del Estudio Jurídico Cotrina y Asociados.



ISBN: 978-9907-9622-0-8

Crossref