



<https://doi.org/10.22199/FIC400586050>



Inteligencia Artificial en la Innovación Social: Guía de Certificación

Desde la Teoría a la Práctica

EDITOR: Emilio Ricci





Ediciones Universidad Católica del Norte

Inteligencia Artificial en la Innovación Social: Guía de Certificación

Desde la Teoría a la Práctica

EDICIONES Universidad Católica del Norte – UCN
Universidad Católica del Norte
Antofagasta – Chile
Domicilio Legal: Av. Angamos 0610, Antofagasta, Chile.

ROR <https://ror.org/02akpm128>

Autores:

Emilio Ricci
José Cassola Bacallao
Brian Keith Norambuena
Pedro Monteros Valdivieso
Diego Llergo Morales
José Marín-Roig Ramón
Janett Guerrero Gallardo

Editor: Emilio Ricci

Primera Edición: Revisada, 2026

Diseño y diagramación: Ediciones Revista Mensaje

Registro de Propiedad Intelectual N.º 2026-A-1123

ISBN: 978-956-287-514-1

© 2026



Las opiniones vertidas en los capítulos de esta obra son de exclusiva responsabilidad de sus autores y no comprometen la postura institucional de la Universidad Católica del Norte ni la de su editor.

La edición de este libro fue financiada por el Gobierno Regional de Antofagasta con recursos de la provisión **Fondo de Innovación para la Competitividad**, FIC-R 2023: "Transferencia Gestión y Certificación del Sello Innovación Social para la Región de Antofagasta", Cód. BIP 40058605-0. Director Prof. Emilio Ricci.



Inteligencia Artificial en la Innovación Social: Guía de Certificación Desde la Teoría a la Práctica


Autores

Emilio Ricci
José Cassola Bacallao
Brian Keith Norambuena
Pedro Monteros Valdivieso
Diego Llergo Morales
José Marín-Roig Ramón
Janett Guerrero Gallardo

Editor

Emilio Ricci

Ediciones Universidad Católica del Norte

 <https://doi.org/10.22199/FIC400586050>



CAPÍTULO V

Ética y Responsabilidad en la IA

Brian Keith Norambuena 

RESUMEN: *La Inteligencia Artificial (IA) no es una tecnología neutral: codifica los valores de quienes la desarrollan y refuerza las estructuras sociales presentes en sus datos de entrenamiento. Partiendo de esta premisa, este capítulo revisa los principios éticos que deberían guiar el desarrollo de estos sistemas, así como los marcos legales que buscan operacionalizarlos, tales como la Recomendación de la UNESCO, los Principios de la OCDE y el Reglamento Europeo de IA. Se examina también el avance regulatorio chileno en la materia y se examinan temas como transparencia algorítmica, privacidad, rendición de cuentas y sesgos, para cerrar con reflexiones sobre el potencial de la IA al servicio del desarrollo sostenible.*

Palabras clave: *IA responsable, regulación de IA, gobernanza algorítmica, desarrollo sostenible.*

ABSTRACT: *Artificial Intelligence (AI) is not neutral technology: it encodes the values of those who develop it and reinforces the social structures present in its training data. Based on this premise, this chapter reviews the ethical principles that should guide the development of these systems, as well as the legal frameworks that seek to operationalize them, such as the UNESCO Recommendation, the OECD Principles, and the European AI Regulation. It also examines Chilean regulatory progress in this area and explores issues such as algorithmic transparency, privacy, accountability, and bias, concluding with reflections on the potential of AI in the service of sustainable development.*

1. Introducción

La inteligencia artificial (IA) pasó de ser una promesa futurista a estar presente en casi todos los aspectos de la vida actual. Desde algoritmos que afinan procesos industriales hasta sistemas que guían decisiones en servicios públicos, la IA define en silencio las oportunidades y limitaciones de millones de personas (O'Neil, 2017; Buolamwini y Gebru, 2018; Eubanks, 2018). Chile ha sido pionero en la regulación de estas tecnologías, con un proyecto de ley presentado en mayo de 2024 (Ministerio de Ciencia, 2024). Sin embargo, el desafío va más allá de los marcos jurídicos: requiere incorporar la ética en las prácticas productivas, los servicios públicos y el desarrollo comunitario. Lejos de ser una limitación, la ética en IA representa una ventaja competitiva: las empresas que abordan estos desafíos generan más confianza y fidelidad del cliente (Capgemini Research Institute, 2019), y la inversión en IA se relaciona con el crecimiento empresarial y la innovación de productos (Babina et al., 2024).

Este capítulo asume una premisa: la IA no es una tecnología neutra, en el sentido definido por Winner en la década de los 80 para los artefactos tecnológicos en general (Winner, 1980). En concreto, los sistemas de IA codifican los valores de sus creadores, los patrones de los datos con los que se entrenan y amplifican las estructuras sociales en las que se despliegan (Crawford, 2021). Dado esto, el desarrollo responsable de la IA no es solo un asunto técnico, sino que plantea dilemas éticos y de gobernanza que exigen una reflexión colectiva y unos marcos normativos adecuados (Mittelstadt et al., 2016). El diseño, la forma en que se implementan y la manera en que se gobiernan estos sistemas definen cómo se reparten las oportunidades y los riesgos en la sociedad (Floridi y COWLS, 2022).

La convergencia de capacidad tecnológica y responsabilidad social hoy requiere respuestas. Mientras el sector productivo adopta optimizaciones con IA e incorpora sistemas autónomos, estas tecnologías abren desafíos como el desplazamiento laboral (Frey y Osborne, 2017; Leopold et al., 2025) y la sostenibilidad ambiental (Strubell et al., 2019; Crawford, 2021). En este capítulo se exploran estas tensiones y se presentan algunos principios éticos clave, el marco regulatorio chileno, los estándares éticos mundiales de la UNESCO, apoyados por 193 países (UNESCO, 2021), y metodologías participativas de diseño de IA centradas en las comunidades afectadas.

2. Marcos regulatorios globales y liderazgo regional

2.1. El liderazgo regulatorio de Chile en América Latina

Chile ha emergido como pionero en gobernanza de IA, posicionándose en la intersección de innovación y responsabilidad. En mayo de 2024, se presentó el Proyecto de Ley N.° 16821-19 para regular los sistemas de IA (UNESCO, 2024; Ministerio de Ciencia, 2024). Se debe notar que, al momento de escribir este capítulo, el proyecto de ley se encuentra aún en trámite legislativo. El marco propuesto adopta un sistema de clasificación basado en riesgos modelado en estándares internacionales, con cuatro categorías (riesgo inaceptable, alto riesgo, riesgo limitado y riesgo mínimo) que reflejan aprendizaje de experiencias internacionales (Ministerio de Ciencia, 2024), de las cuales se describen a continuación el riesgo inaceptable y el alto riesgo, al ser las más relevantes para el contexto de este capítulo.

La norma propuesta (Alessandri Abogados, 2024) prohíbe los sistemas de IA de riesgo inaceptable, como los de manipulación subliminal, los que aprovechan la vulnerabilidad por edad o discapacidad, los de categorización biométrica de datos sensibles (como las opiniones políticas o la religión) y los de puntuación social por parte de las autoridades. Los sistemas de IA de alto riesgo, es decir, aquellos que impactan en la salud, seguridad, derechos fundamentales o protección al consumidor, están sujetos a obligaciones estrictas inspiradas por casos internacionales (Buolamwini y Gebru, 2018; Angwin y Larson, 2022). Estas obligaciones incluyen evaluaciones de riesgo obligatorias, protocolos de gobernanza de datos, transparencia y supervisión humana.

La arquitectura de gobernanza propuesta por la ley establece un Consejo Técnico Asesor sobre IA que asiste al Ministerio de Ciencia, con potenciales funciones de acreditación y auditoría, y una futura Agencia de Protección de Datos Personales encargada de la supervisión del cumplimiento de la normativa (BCN, 2024). Además, el marco promueve la innovación con *sandboxes* regulatorios (Ministerio de Economía, 2021), entornos protegidos para experimentar con sistemas de IA en condiciones reales antes de su despliegue (Ministerio de Ciencia, 2024).

Por otro lado, se debe destacar la Política Nacional de Inteligencia Artificial (Ministerio de Ciencia, 2021), basada en tres ejes: factores habilitantes, desarrollo y adopción, y gobernanza y ética. El documento presenta una visión a largo plazo (hasta 2031) con orientaciones para fortalecer las competencias ciudadanas en IA y generar discusión pública sobre los efectos legales, éticos, sociales y económicos de estas tecnologías. Esta política fue actualizada en 2024 para abordar los desafíos emergentes, con foco en gobernanza y ética, siendo Chile el primer país en implementar la Metodología de Evaluación del Estado de Preparación (RAM, por sus siglas en inglés) de la UNESCO, cuyos resultados informaron la actualización de la política (UNESCO, 2023).

2.2. Marcos éticos y regulatorios internacionales

La ley chilena se alinea con tres marcos internacionales que definen la ética de la IA a nivel mundial. El primero es la Recomendación de la UNESCO sobre la ética de la IA, adoptada en 2021 por los 193 Estados miembros, el primer estándar global sobre ética en IA (UNESCO, 2021). Sus cuatro valores centrales son el respeto por los derechos humanos y dignidad de las personas, la vida en sociedades pacíficas y justas, el aseguramiento de la diversidad e inclusión, y el florecimiento ambiental.

Los principios declarados por la UNESCO proporcionan orientación concreta para operacionalizar estos valores (UNESCO, 2021). El principio de proporcionalidad y no hacer daño requiere que el despliegue de IA se limite a lo justo y necesario para sus objetivos, con evaluación de riesgos para prevenir daños. El principio de seguridad aborda tanto peligros físicos como vulnerabilidades de ciberseguridad. El principio de equidad y no discriminación requiere que los sistemas de IA promuevan la justicia social y eviten resultados sesgados. El principio de sostenibilidad exige evaluar los impactos sociales, culturales, económicos y ambientales de la IA. El principio de privacidad y protección de datos demanda marcos adecuados de protección a lo largo del ciclo de vida de la IA. El principio de supervisión y determinación humanas establece que los seres humanos deben mantener el control y la responsabilidad última sobre los sistemas de IA. El principio de transparencia y explicabilidad requiere que los sistemas de IA sean auditables y trazables, con mecanismos de supervisión contextualizados. El principio de responsabilidad y rendición de cuentas exige que se pueda atribuir responsabilidad ética y legal a humanos. El princi-

pio de conciencia y alfabetización promueve la comprensión pública de la IA y los datos. Finalmente, el principio de gobernanza multipartípite y colaboración adaptativa requiere que la gobernanza de IA sea inclusiva y flexible ante nuevos desarrollos tecnológicos.

El segundo corresponde a los Principios de IA de la OCDE, originalmente adoptados en 2019 y actualizados en 2024, que establecen estándares complementarios respaldados por la mayoría de los gobiernos (OCDE, 2024). Estos establecen cinco pilares para una IA confiable: (1) promover el crecimiento inclusivo, la sostenibilidad y el bienestar; (2) situar a las personas en el centro con criterios de equidad; (3) asegurar transparencia y capacidad de explicación; (4) garantizar sistemas robustos, seguros y protegidos; y (5) establecer mecanismos de rendición de cuentas. Estos pilares convergen con los de la UNESCO, reforzando la existencia de un consenso internacional emergente sobre los fundamentos éticos de la IA.

Finalmente, el tercero corresponde al Reglamento Europeo de IA, que proporciona la primera regulación de IA legalmente vinculante del mundo (Comisión Europea, 2024). Aunque Chile no es territorio de la Unión Europea, los efectos extraterritoriales del Reglamento y su poder de configuración del mercado lo hacen globalmente influyente: existe el denominado “efecto Bruselas”, por el cual las regulaciones europeas se convierten en estándares internacionales de facto (Bradford, 2020; GPRG, 2024). El sistema de clasificación basado en riesgos del Reglamento Europeo influyó directamente en el diseño legislativo chileno (White & Case, 2024).

3. Principios éticos fundamentales en Inteligencia Artificial

Los marcos internacionales descritos en la sección anterior proporcionan orientaciones de política, pero traducir estos principios a prácticas organizacionales concretas requiere herramientas adicionales. En esta sección articulamos algunos de los principios de la UNESCO usando la estructura de la ética biomédica clásica, que organiza las consideraciones éticas en torno a beneficencia, no maleficencia, autonomía y justicia. Esta aproximación permite relacionar los principios internacionales con un marco conceptual establecido: la beneficencia se vincula con los valores de dignidad humana y mejora de la calidad de vida, en tanto ambos orientan la IA hacia el bienestar integral de personas y entornos; la no maleficencia corresponde al principio de proporcionalidad y “no hacer daño”; la autonomía se relaciona con el principio de supervisión y determinación humanas; y la justicia se alinea con los principios de equidad y no discriminación. Esta correspondencia no es exacta ni exhaustiva, pero su amplio reconocimiento en contextos de decisiones que afectan a personas facilita el diálogo interdisciplinario y la aplicación práctica.

3.1. Beneficencia y orientación al bien común

El principio de beneficencia establece que los sistemas de IA deben diseñarse e implementarse con el propósito explícito de generar beneficios para las personas y la sociedad. Para hacer efectivo este principio, los creadores de sistemas de IA deben definir desde el principio los beneficiarios del sistema y cómo se repartirán los beneficios que genere (UNESCO, 2021). Un sistema de IA que optimiza procesos productivos cumple con el principio de beneficencia cuando sus beneficios se distribuyen deliberadamente: por ejemplo, cuando las ganancias de eficiencia se reinvierten en mejores condiciones laborales, capacitaciones o creación de nuevos roles de mayor valor agregado para los trabajadores (Didier, 2023).

La maximización de ganancias económicas sin considerar a los beneficiarios no satisface este principio.

3.2. No maleficencia y gestión de riesgos

El principio de no maleficencia establece el imperativo de evitar causar daño. Entre los daños posibles que pueden causar los sistemas de IA están los sesgos algorítmicos, las violaciones de la privacidad, el desplazamiento laboral y la concentración de poder. La UNESCO expresa esto con el principio de proporcionalidad, el cual exige que se realice una evaluación de riesgos para prevenir daños potenciales, incluidos los efectos directos, indirectos y acumulativos (UNESCO, 2021). Por ejemplo, si bien la automatización ha hecho más segura la operación productiva, estas mismas tecnologías que protegen a los trabajadores pueden sustituirlos, por lo que esto debe ser evaluado al momento de diseñar un sistema de este tipo (Didier, 2023).

3.3. Autonomía y supervisión humana

El respeto por la autonomía humana significa que las personas deben mantener el control sobre las decisiones de su vida sin manipulación algorítmica. Este principio exige que las IA sean extensiones de la agencia humana, no sustitutos de ella. Según la UNESCO, los humanos deben seguir siendo responsables y no puede haber una delegación de rendición de cuentas a los algoritmos (UNESCO, 2021). De este principio se deriva que, independientemente del nivel de automatización, siempre debe existir un ser humano identificable que asuma la responsabilidad de las decisiones (Raji *et al.*, 2020). Esta supervisión debe ser significativa: las personas a cargo deben tener la información, el tiempo y la autoridad para auditar e intervenir en las decisiones algorítmicas cuando sea necesario.

3.4. Justicia, equidad y no discriminación

La justicia en IA implica la distribución justa de beneficios y cargas y la prevención de sesgos algorítmicos. Los modelos de IA pueden reproducir y amplificar los sesgos históricos en los datos, lo que lleva a tratamientos desiguales para ciertos grupos. Estos sesgos a menudo actúan de manera sutil, imperceptible, pero con impacto en las personas afectadas (UNESCO, 2021). Se debe notar que la equidad algorítmica no tiene una única definición: puede ser igualdad de oportunidades de resultado positivo entre grupos, igualdad de precisión del sistema para todos los grupos o igualdad de tratamiento para personas similares. Cada definición conlleva ciertos valores, y la elección de la más apropiada dependerá del contexto y sus consecuencias prácticas al implementar el sistema (OCDE, 2024).

4. Transparencia y explicabilidad

4.1. El imperativo de la transparencia algorítmica

La transparencia es un requisito para el desarrollo responsable de la IA. Sin conocimiento de cómo funcionan los algoritmos, no se puede evaluar adecuadamente su nivel de justicia, descubrir posibles sesgos ni crear formas efectivas de rendición de cuentas. El Regla-

mento Europeo de IA impone deberes de transparencia para los sistemas de alto riesgo (Comisión Europea, 2024), en un reconocimiento de que la caja negra algorítmica es un problema para la gobernanza democrática de la tecnología.

La transparencia en IA opera en múltiples niveles (Selbst y Barocas, 2018). A nivel técnico, implica la documentación de los datos de entrenamiento, las características del modelo, las métricas de rendimiento y las limitaciones conocidas. A nivel organizacional, requiere claridad sobre los procesos de toma de decisiones, las personas responsables y los mecanismos de supervisión. A nivel social, demanda comunicación accesible que permita a los afectados por decisiones algorítmicas comprender las razones de dichas decisiones. La tensión existente entre la transparencia e intereses propietarios requiere una resolución equilibrada; mientras las organizaciones legítimamente protegen innovaciones, la transparencia fundamental sobre sistemas de IA que afectan derechos no puede subordinarse a la confidencialidad comercial.

4.2. Herramientas para la explicabilidad

La IA explicable ha evolucionado de aspiración teórica a necesidad práctica con respaldo regulatorio. Existen herramientas que permiten entender cómo un sistema de IA llegó a una decisión particular (Lundberg y Lee, 2017; Ribeiro *et al.*, 2016). Algunas calculan cuánto contribuyó cada factor a la decisión final, permitiendo identificar si ciertos factores tienen una influencia inapropiada. Otras crean explicaciones simplificadas de decisiones específicas a través de visualizaciones comprensibles para personas sin formación técnica.

No obstante, si bien las herramientas técnicas de explicabilidad son necesarias, resultan insuficientes para asegurar la transparencia real, que requiere comunicación contextualizada para diferentes grupos de interés. Por ejemplo, cuando un algoritmo determina la asignación de recursos hídricos, las explicaciones deben satisfacer a diversos interesados: los operadores que buscan métricas de eficiencia, las autoridades regulatorias que evalúan cumplimiento legal, los líderes comunitarios que evalúan impacto local y los científicos ambientales que monitorean sostenibilidad ecológica. Las mejores prácticas incluyen documentación estructurada que proporciona información sobre capacidades, limitaciones, usos previstos y aplicaciones inapropiadas de los sistemas de IA (Mitchell *et al.*, 2019; Gebru *et al.*, 2021).

5. Privacidad y protección de datos personales

5.1. El derecho a la privacidad en la era algorítmica

Los sistemas de IA contemporáneos dependen de grandes volúmenes de datos para su entrenamiento, y su capacidad de inferencia permite extraer información sensible incluso de datos aparentemente inofensivos (Acquisti *et al.*, 2016; Zuboff, 2023). Las operaciones productivas modernas intensifican este fenómeno al generar datos sin precedentes a través de sensores, telemetría de vehículos, monitoreo de trabajadores, seguimiento de rendimiento de equipos y mediciones ambientales. Esta recolección masiva plantea interrogantes sobre el almacenamiento y el mal uso de información personal sensible.

Frente a estos desafíos, Chile ha avanzado en la modernización de su marco normativo. La reforma de la Ley N.º19.628, cuyo texto final fue aprobado por comisión mixta

en julio de 2024, alinea la protección de datos con los estándares internacionales más exigentes. La legislación actualizada introduce ocho principios fundamentales: licitud, limitación de finalidad, proporcionalidad, calidad de datos, responsabilidad, seguridad, transparencia y confidencialidad. También establece derechos individuales ampliados, incluyendo la portabilidad de datos, y define la creación de una futura Agencia de Protección de Datos Personales independiente con jurisdicción sobre cualquier entidad que ofrezca servicios a residentes chilenos, independientemente de su ubicación física (Carey, 2024). Se debe notar que al momento de escribir este capítulo la ley no se encuentra plenamente implementada, pues la Agencia aún no se encuentra en funcionamiento.

5.2. Privacidad por diseño y tecnologías de protección

El concepto de privacidad por diseño (Cavoukian, 2009), incorporado en regulaciones internacionales y en la reforma chilena de protección de datos, establece que la protección de la privacidad debe integrarse desde las etapas más tempranas del diseño de sistemas. Los siete principios fundacionales incluyen: protección proactiva, privacidad como configuración predeterminada, privacidad integrada en el diseño, funcionalidad completa sin falsas disyuntivas, seguridad de extremo a extremo, visibilidad y transparencia, y respeto por el usuario.

Existen técnicas que permiten analizar datos de manera agregada sin exponer información individual. La privacidad diferencial (Dwork y Roth, 2014) es una técnica que consiste en añadir pequeñas cantidades de ruido aleatorio a los datos, de manera que los patrones generales permanecen visibles para el análisis, pero resulta imposible identificar a individuos específicos. Este enfoque ha sido aplicado a gran escala, incluyendo su uso en censos nacionales y sistemas de telemetría. Por otro lado, el aprendizaje federado (McMahan *et al.*, 2017) es otra técnica relevante que permite entrenar modelos de IA sobre datos distribuidos sin centralizar la información, reduciendo así las vulnerabilidades frente a brechas de seguridad.

5.3. Soberanía de datos y comunidades locales

La privacidad en la IA es una extensión de la soberanía de datos de las comunidades locales. La Declaración de las Naciones Unidas sobre los Derechos de los Pueblos Indígenas (ONU, 2007) reconoce el derecho de los pueblos indígenas a conservar, proteger y desarrollar su patrimonio cultural, sus conocimientos tradicionales y sus expresiones culturales, y a determinar las prioridades y estrategias para el uso de los recursos en sus territorios.

Este marco se extiende a los datos como recurso: la información sobre territorios, prácticas tradicionales, conocimiento ecológico y características comunitarias queda comprendida dentro de la soberanía de datos. Por lo tanto, cuando sistemas de IA recolectan datos en territorios comunitarios (ya sea mediante monitoreo ambiental, análisis satelital o aplicaciones de gestión), las comunidades deberían decidir cómo se recolectan, almacenan, analizan y comparten. En la práctica, esto implica obtener consentimiento informado previo, comunicar los alcances del proyecto de manera accesible, establecer acuerdos de distribución de beneficios cuando los datos generen valor y respetar las decisiones de restringir el acceso externo.

6. Responsabilidad y rendición de cuentas

6.1. El problema de la atribución de responsabilidad

La pregunta sobre quién es responsable cuando un sistema de IA produce daños ha dejado de ser una abstracción filosófica (Raji *et al.*, 2020; Selbst, 2021). Los sistemas de IA introducen complejidades particulares para la atribución de responsabilidad. La cadena de desarrollo involucra múltiples actores: diseñadores de algoritmos, proveedores de datos, integradores de sistemas, operadores y usuarios finales. El problema radica en que cada uno puede contribuir al daño, pero ninguno controla el sistema completo, dificultando atribuir responsabilidad a un actor específico. Los marcos de auditoría algorítmica proporcionan metodologías para abordar este desafío (Koshiyama *et al.*, 2024), abarcando etapas que incluyen: configuración de datos y tareas, preprocesamiento de características, selección de modelos, posprocesamiento y reporte, y despliegue en producción.

6.2. Gobernanza organizacional de la IA

Los requisitos de rendición de cuentas del Reglamento Europeo de IA establecen obligaciones para sistemas de alto riesgo, incluyendo gestión de riesgos a lo largo del ciclo de vida, evaluaciones de conformidad, monitoreo posterior al despliegue y reporte de incidentes serios (Comisión Europea, 2024). El marco propuesto por Chile adopta requisitos similares adaptados al contexto nacional. Para las organizaciones, esto se traduce en establecer comités de gobernanza de IA con representación de equipos técnicos, cumplimiento legal, especialistas en ética, representantes de trabajadores y partes interesadas comunitarias. Estos comités deberían revisar despliegues de IA antes del lanzamiento, monitorear métricas de rendimiento, investigar incidentes prontamente y autorizar modificaciones cuando surjan problemas.

La auditoría por terceros, aunque genera costos, proporciona credibilidad que la auto-certificación no puede igualar. Este modelo, análogo al de la auditoría financiera donde las organizaciones contratan auditores independientes que enfrentan responsabilidad profesional por revisiones inadecuadas, ya ha comenzado a materializarse en el ámbito de la IA con estándares como ISO/IEC 42001:2023, que exige verificación independiente (ISO/IEC 2023). El Consejo Técnico Asesor sobre IA podría acreditar firmas auditoras, establecer estándares y metodologías de auditoría, y mantener registros de evaluaciones de sistemas de alto riesgo (BCN, 2024).

7. Herramientas prácticas para la evaluación ética

7.1. Estándares y certificaciones internacionales

La norma ISO/IEC 42001:2023, publicada en diciembre de 2023 y el primer estándar certificable de sistemas de gestión de IA en el mundo, especifica los requisitos para establecer, implementar, mantener y mejorar sistemas de gestión de IA. El estándar trata aspectos específicos de la IA, como ética, transparencia y aprendizaje continuo, y exige a las organizaciones definir procesos de gestión de riesgos, evaluación de impacto de sistemas de IA, gestión del ciclo de vida, seguimiento de proveedores externos y demostrar una IA responsable y confiable (ISO/IEC, 2023). La certificación se audita y verifica de forma

independiente por terceros y tiene una validez de 3 años, con vigilancia anual y exigencias de recertificación. En ese sentido, el estándar unifica marcos de gestión ya existentes (seguridad de la información, gestión de la calidad, etc.) y es compatible con normativas internacionales de protección de datos, el Marco de IA de la OCDE y otros marcos de gestión de riesgos de IA.

La herramienta de evaluación ALTAI (Assessment List for Trustworthy AI), desarrollada por el Grupo de Expertos de Alto Nivel sobre IA de la Unión Europea, proporciona una herramienta gratuita de autoevaluación organizada en torno a siete requisitos: (1) supervisión y control humano, (2) solidez técnica y seguridad, (3) protección de datos y su gobernanza, (4) transparencia, (5) equidad, diversidad y no discriminación, (6) impacto positivo en la sociedad y el medio ambiente, y (7) responsabilidad (Comisión Europea, 2024), permitiendo a las organizaciones evaluar los sistemas de IA contra criterios de confiabilidad, identificar brechas y priorizar mejoras.

7.2. Detección y mitigación de sesgos

Existen herramientas que permiten detectar y corregir sesgos en sistemas de IA (Bellamy *et al.*, 2019; Weerts *et al.*, 2023). Estas herramientas funcionan con métricas que miden diferentes aspectos de la equidad, por ejemplo, si diferentes grupos reciben resultados similares, si el sistema tiene la misma precisión o si comete los mismos tipos de errores independientemente del grupo. Una vez identificados los sesgos, las herramientas ofrecen métodos para corregirlos.

Las estrategias de corrección operan en tres momentos (Mehrabi *et al.*, 2021). Antes de entrenar el sistema, se pueden ajustar los datos para asegurar que todos los grupos estén adecuadamente representados y que no se reproduzcan patrones históricos de discriminación. Durante el entrenamiento, se pueden incorporar restricciones que obliguen al sistema a tratar equitativamente a diferentes grupos. Después del entrenamiento, se pueden ajustar los resultados del sistema para equilibrar su comportamiento entre grupos.

7.3. Diseño participativo con comunidades afectadas

Los marcos de ética técnica más sofisticados fracasan cuando excluyen a quienes serán más afectados por los sistemas de IA (Birhane *et al.*, 2022; Delgado *et al.*, 2023). En particular, el diseño participativo trata a los usuarios como expertos en su propia experiencia y colaboradores iguales a lo largo de los procesos de diseño. Para sistemas de IA que afectan a comunidades, esto significa involucrar a trabajadores, residentes y otros grupos de interés no meramente a través de consulta, sino a través de codiseño con autoridad compartida en la toma de decisiones.

Arnstein (1969) propuso una escala de ocho niveles de participación ciudadana, desde la manipulación hasta el control ciudadano. Adaptando este marco al contexto de IA, se distinguen cuatro niveles (OCDE, 2024): consulta, donde se proporciona *input* fuera del desarrollo central sin garantía de impacto en el diseño; contribución, donde los participantes proporcionan *input* que informa el desarrollo; colaboración, donde existe asociación continua a lo largo del desarrollo; y codiseño, donde existe asociación plena con toma de decisiones compartida.

8. IA como motor de Innovación Social

8.1. IA responsable: del imperativo ético a la ventaja competitiva

La IA ética es, ante todo, un imperativo normativo derivado del respeto a la dignidad humana y los derechos fundamentales. Sin embargo, la evidencia empírica muestra que este imperativo y los intereses organizacionales no están en conflicto: la IA responsable se reconoce también como ventaja competitiva (Capgemini Research Institute, 2019; Zhang *et al.*, 2021; Accenture, 2022). En este contexto, las organizaciones que invierten en ética de IA logran mayores beneficios operativos de sus implementaciones que aquellas que no lo hacen. Este diferencial de rendimiento refleja múltiples mecanismos: la confianza se convierte en valor a través de la fidelidad de clientes y usuarios; la reputación ética atrae inversores y talento; la ética proactiva reduce sanciones regulatorias y costos de litigio; y el diseño inclusivo expande los mercados.

Para organizaciones que operan en contextos donde dependen de la aceptación comunitaria para sus operaciones, demostrar que los despliegues de IA benefician en lugar de perjudicar a las poblaciones locales resulta fundamental (Moffat y Zhang, 2014; Candelon *et al.*, 2023). Cuando las comunidades perciben que los despliegues de IA mejoran la seguridad en lugar de meramente desplazar empleo, o mejoran el monitoreo ambiental en lugar de obscurecer impactos, es más probable que apoyen las operaciones continuadas.

El rendimiento en criterios Ambientales, Sociales y de Gobernanza influencia fuertemente las decisiones de inversión (Eccles *et al.*, 2014). Los principales inversores institucionales utilizan estas calificaciones para evaluar sus participaciones. En este contexto, las organizaciones que demuestran gobernanza de IA a través de certificaciones, informes transparentes e involucramiento de partes interesadas logran mejores calificaciones, pudiendo acceder a capital a menor costo y atrayendo inversores alineados con valores de sostenibilidad.

8.2. IA para los Objetivos de Desarrollo Sostenible

Los 17 ODS de las Naciones Unidas proporcionan un marco que conecta las capacidades de IA con las prioridades de desarrollo (Covels *et al.*, 2021).

En relación con el ODS 3 sobre salud y bienestar, la IA puede contribuir a expandir el acceso a servicios de salud en áreas remotas donde existe escasez de especialistas (Topol, 2019). Los sistemas de diagnóstico pueden ayudar a profesionales de salud a identificar condiciones que requieren atención especializada. Los requisitos éticos (OMS, 2024) incluyen asegurar que estos sistemas tengan precisión equivalente para diferentes poblaciones, que complementen en lugar de reemplazar el juicio clínico humano, y que protejan la privacidad de los datos.

Respecto al ODS 6 sobre agua limpia y saneamiento, la IA puede contribuir a optimizar la gestión de recursos hídricos escasos (Garrido-Baserba *et al.*, 2020). Los sistemas inteligentes pueden detectar fugas en infraestructura, predecir demanda, y apoyar decisiones de asignación. Sin embargo, los requisitos éticos son exigentes: los sistemas deben incorporar criterios de equidad, no solo de eficiencia; deben establecer protecciones para usos prioritarios como el consumo humano; y las decisiones importantes deben estar sujetas a supervisión humana y participación comunitaria.

En cuanto al ODS 8 sobre trabajo decente y crecimiento económico, los impactos de la automatización sobre el empleo demandan programas proactivos de transición laboral (Santoni de Sio, 2024). La IA puede apoyar esto a través de análisis de habilidades que identifiquen capacidades de la fuerza laboral y necesidades de capacitación, plataformas de vinculación que conecten personas con oportunidades, y sistemas de aprendizaje adaptativo para recapacitación. Los requisitos éticos incluyen asegurar que quienes son afectados por la automatización reciban apoyo genuino, que participen en el diseño de los programas de transición y que los beneficios de la mayor productividad se compartan equitativamente.

Respecto al ODS 13 sobre acción por el clima, la IA puede contribuir a reducir el impacto ambiental de las actividades productivas (Rolnick *et al.*, 2022). Los sistemas inteligentes pueden optimizar el uso de energía, predecir necesidades de mantenimiento para prevenir ineficiencias, y monitorear indicadores. Los requisitos éticos incluyen transparencia en el reporte de resultados ambientales, acceso de las comunidades a información sobre impactos locales, y mecanismos para que las comunidades puedan cuestionar datos o decisiones que consideren problemáticos.

8.3. Inteligencia Artificial al servicio del desarrollo social

La IA está transformando la manera en que resolvemos problemas sociales complejos, desde mejorar los servicios públicos hasta predecir crisis humanitarias. Sin embargo, aprovechar este potencial requiere orientar explícitamente el desarrollo de IA hacia el bienestar colectivo, más allá de maximizar la eficiencia operacional o los beneficios económicos (Tomašev *et al.*, 2020; Stahl, 2021). En este contexto, el concepto de “IA para el bien social” (AI for Social Good) ha emergido como un campo que busca aplicar técnicas de aprendizaje automático y análisis de datos a problemas de interés público (Tomašev *et al.*, 2020; Cowsls *et al.*, 2021; Floridi *et al.*, 2021).

El desarrollo de la IA para el bien social enfrenta desafíos como la falta de datos de alta calidad en los países en vías de desarrollo, la falta de infraestructura informática en los lugares vulnerables y la desconexión entre los creadores de IA y las comunidades que podrían beneficiarse de la esta (Tomašev *et al.*, 2020). Para abordar estos retos, se necesitan modelos colaborativos con organizaciones de la sociedad civil, gobiernos locales y comunidades afectadas desde el principio. Para que la IA realmente sirva como motor de desarrollo e IS, las empresas deben incorporar la ética en sus modelos de negocio. Esto implica una gobernanza orientada al beneficio social y protección de los trabajadores, participación comunitaria, sostenibilidad ambiental, respeto a la privacidad y la dignidad humana. Estos factores facilitan la creación de confianza, legitimidad y apoyo de las diferentes partes interesadas (Stahl, 2021).

Conclusiones

El futuro de la IA en Chile será determinado no solo por las capacidades tecnológicas, sino por los valores incorporados en los algoritmos, las estructuras de gobernanza que rodean su implementación, y la distribución de beneficios y cargas a través de la sociedad. El liderazgo regulatorio de Chile se ha materializado en los avances de propuestas legislativas. Asimismo, existen múltiples marcos prácticos y estándares que pueden guiar la imple-

mentación y adopción de sistemas de IA. No obstante, la legislación y herramientas por sí solos no pueden asegurar la innovación responsable. La implementación efectiva requiere el compromiso activo de todos los actores involucrados: organizaciones que desarrollan y despliegan sistemas de IA, trabajadores que interactúan con estos sistemas, comunidades que experimentan sus impactos y autoridades que supervisan su uso.

A nivel local, se debe invertir en la capacidad y alfabetización regional de IA, asegurando que las regiones desarrollen talento local en lugar de dependencia de consultores externos. Esto incluye programas formativos que capaciten especialistas en IA con integración de ética, educación comunitaria que permita a los ciudadanos comprender y cuestionar sistemas algorítmicos, capacitación de trabajadores que prepare a los empleados para roles en evolución y desarrollo de capacidades comunitarias que apoye la participación informada en decisiones tecnológicas. La transferencia tecnológica sin desarrollo de capacidades locales perpetúa patrones donde las regiones proporcionan recursos mientras la captura de valor ocurre en otros lugares.

La pregunta fundamental no es tecnológica, sino de gobernanza: ¿a qué valores servirá la IA, y quién decidirá? Los marcos regulatorios proporcionan fundamentos, los estándares internacionales ofrecen orientación y las herramientas prácticas permiten la implementación. No obstante, la clave está en la voluntad colectiva de asegurar que la IA sirva a la innovación para las personas en lugar de meramente a la innovación en procesos. El desafío y la oportunidad están definidos; la responsabilidad de actuar recae en todos quienes participan en el ecosistema de IA de Chile.

Referencias bibliográficas

- Accenture. (2022). The art of AI maturity: Advancing from practice to performance. Disponible en: <https://www.accenture.com/us-en/insights/artificial-intelligence/ai-maturity-and-transformation> (Acceso: 30-11-25)
- Acquisti, A., Taylor, C., & Wagman, L. (2016). The economics of privacy. *Journal of Economic Literature*, 54(2), 442-492.
- Alessandri Abogados. (2024). Proyecto de ley que regula la inteligencia artificial ingresa al Congreso. Disponible en: <https://alessandri.legal/proyecto-de-ley-que-regula-la-inteligencia-artificial-ingreso-al-congreso/> (Acceso: 30-11-25)
- Angwin, J., & Larson, J. (2022). Bias in criminal risk scores is mathematically inevitable, researchers say. In *Ethics of data and analytics* (pp. 265-267). Auerbach Publications.
- Arnstein, S. R. (1969). A ladder of citizen participation. *Journal of the American Institute of Planners*, 35(4), 216-224.
- Asamblea General de las Naciones Unidas (ONU). (2007). Declaración sobre los derechos de los pueblos indígenas. Resolución 61/295. Disponible en: https://www.un.org/es/events/indigenousday/pdf/indigenousdeclaration_faqs.pdf (Acceso: 30-11-25)
- Babina, T., Fedyk, A., He, A., & Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, 151, 103745.
- Bellamy, R. K. E., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., Lohia, P., Martino, J., Mehta, S., Mojsilović, A., Nagar, S., Ramamurthy, K. N., Richards, J., Saha, D., Sattigeri, P., Singh, M., Varshney, K. R., & Zhang, Y. (2019). AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research and Development*, 63(4/5), 4:1-4:15.

- Biblioteca del Congreso Nacional de Chile (BCN). (2024). Resultado consulta experta IA. Disponible en: https://obtienearchivo.bcn.cl/obtienearchivo?id=repositorio/10221/36220/2/Resultado_Consulta_experta_IA____BCN.pdf (Acceso: 30-11-25)
- Birhane, A., Isaac, W., Prabhakaran, V., Diaz, M., Elish, M. C., Gabriel, I., & Mohamed, S. (2022). Power to the people? Opportunities and challenges for participatory AI. En *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (pp. 1-8).
- Bradford, A. (2020). *The Brussels effect: How the European Union rules the world*. Oxford University Press.
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*. PMLR.
- Candelon, F., di Carlo, R. C., & Mills, S. (2023). *Why AI Needs a Social License*. BCG Global, February, 22.
- Capgemini Research Institute. (2019). Why addressing ethical questions in AI will benefit organizations. Disponible en: https://www.capgemini.com/gb-en/wp-content/uploads/sites/5/2022/05/AI-in-Ethics_Web.pdf (Acceso: 30-11-25)
- Carey. (2024). Comisión mixta aprueba el texto final del proyecto que modifica la Ley N.º 19.628 sobre protección de la vida privada. Disponible en: <https://www.carey.cl/comision-mixta-aprueba-el-texto-final-del-proyecto-de-ley-que-modifica-la-ley-19-628-sobre-proteccion-de-la-vida-privada> (Acceso: 30-11-25)
- Cavoukian, A. (2009). Privacy by design: The 7 foundational principles. *Information and privacy commissioner of Ontario, Canada*, 5(2009), 12.
- Comisión Europea. (2024). Marco regulatorio para la inteligencia artificial (AI Act). Disponible en: <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> (Acceso: 30-11-25)
- Cowls, J., Tsamados, A., Taddeo, M., & Floridi, L. (2021). A definition, benchmark and database of AI for social good initiatives. *Nature Machine Intelligence*, 3(2), 111-115.
- Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- Delgado, F., Yang, S., Madaio, M., & Yang, Q. (2023). The participatory turn in ai design: Theoretical foundations and the current state of practice. En *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*.
- Didier, N. (2023). *Just labor transitions: Workforce policies in times of technological change* (Tesis de Doctorado, Arizona State University).
- Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and trends® in theoretical computer science*, 9(3-4), 211-407.
- Eccles, R. G., Ioannou, I., & Serafeim, G. (2014). The impact of corporate sustainability on organizational processes and performance. *Management science*, 60(11), 2835-2857.
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. Macmillan+ ORM.
- Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. *Machine learning and the city: Applications in architecture and urban design*, 535-545.
- Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2021). How to design AI for social good: Seven essential factors. In *Ethics, governance, and policies in artificial intelligence* (pp. 125-151). Cham: Springer International Publishing.

- Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation?. *Technological forecasting and social change*, 114, 254-280.
- Garrido-Baserba, M., Corominas, L., Cortés, U., Rosso, D., & Poch, M. (2020). The fourth-revolution in the water sector encounters the digital revolution. *Environmental Science & Technology*, 54(8).
- Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12), 86-92.
- Global Policy Regulation Group (GPRG). (2024). La geopolítica de la regulación de IA. Disponible en: <https://www.globalprg.org/publications/the-geopolitics-of-ai-regulation> (Acceso: 30-11-25)
- ISO/IEC. (2023). ISO/IEC 42001:2023 - Tecnología de la información – Inteligencia artificial – Sistema de gestión. Disponible en: <https://www.iso.org/standard/81230.html> (Acceso: 30-11-25)
- Koshiyama, A., Kazim, E., Treleaven, P., Rai, P., Szpruch, L., Pavey, G., Ahamat, G., Leutner, F., Goebel, R., Knight, A., Adams, J., Hitrova, C., Barnett, J., Nachev, P., Barber, D., Chamorro-Premuzic, T., Klemmer, K., Gregorovic, M., Khan, S., Lomas, E., Hilliard, A., & Chatterjee, S. (2024). Towards algorithm auditing: Managing legal, ethical and technological risks of AI, ML and associated algorithms. *Royal Society Open Science*, 11(5).
- Leopold, T., Di Battista, A., Jativa, X., Sharma, S., Li, R., & Grayling, S. (2025). Future of jobs report 2025. In *World Economic Forum*.
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics* (pp. 1273-1282). PMLR.
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, 54(6), 1-35.
- Ministerio de Ciencia, Tecnología, Conocimiento e Innovación. (2021). Política Nacional de Inteligencia Artificial. Biblioteca del Congreso Nacional de Chile. <https://www.bcn.cl/leychile/navegar?idNorma=1169399> (Acceso: 30-11-25)
- Ministerio de Ciencia, Tecnología, Conocimiento e Innovación de Chile. (2024). Proyecto de Ley que regula los sistemas de IA. Disponible en: <https://www.minciencia.gob.cl/areas/inteligencia-artificial/Inteligencia-Artificial/Proyecto-Ley-regula-sistemas-IA/> (Acceso: 30-11-25)
- Ministerio de Economía, Fomento y Turismo. (2021). Sandbox Regulatorio de Inteligencia Artificial en Chile: Documento para Discusión. Disponible en: <https://www.economia.gob.cl/wp-content/uploads/2021/09/PaperSandboxIA.pdf> (Acceso: 30-11-25).
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. En *Proceedings of the Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 2053951716679679.
- Moffat, K., & Zhang, A. (2014). The paths to social licence to operate: An integrative model explaining community acceptance of mining. *Resources Policy*, 39, 61-70.

- OCDE. (2024). Principios de IA de la OCDE. Disponible en: <https://oecd.ai/en/ai-principles> (Acceso: 30-11-25)
- O’Neil, C. (2017). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown.
- Organización Mundial de la Salud (OMS). (2024). Ethics and governance of artificial intelligence for health: large multi-modal models. WHO guidance.
- 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. Association for Computing Machinery.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?” Explaining the predictions of any classifier. En Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).
- Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A. S., Maharaj, T., Sherwin, E. D., Mukkavilli, S. K., Kording, K. P., Gomes, C. P., Ng, A. Y., Hassabis, D., Platt, J. C., Creutzig, F., Chayes, J., & Bengio, Y. (2022). Tackling climate change with machine learning. *ACM Computing Surveys*, 55(2), Article 42.
- Santoni de Sio, F. (2024). Artificial intelligence and the future of work: Mapping the ethical issues. *The Journal of Ethics*, 28(3), 407-427.
- Selbst, A. D. (2021). An institutional view of algorithmic impact assessments. *Harv. JL & Tech.*, 35, 117.
- Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. *Fordham L. Rev.*, 87, 1085.
- Stahl, B. C. (2021). Artificial intelligence for a better future: an ecosystem perspective on the ethics of AI and emerging digital technologies (p. 124). Springer Nature.
- Strubell, E., Ganesh, A., & McCallum, A. (2019, July). Energy and policy considerations for deep learning in NLP. En Proceedings of the 57th annual meeting of the association for computational linguistics.
- Tomašev, N., Cornebise, J., Hutter, F., Mohamed, S., Picciariello, A., Connelly, B., Belgrave, D. C. M., Ezer, D., Haert, F. C. van der, Mugisha, F., Abila, G., Arai, H., Almiraat, H., Proskurnia, J., Snyder, K., Otake-Matsuura, M., Othman, M., Glasmachers, T., Wever, W. de, Teh, Y. W., Khan, M. E., Winne, R. D., Schaul, T., & Clopath, C. (2020). AI for social good: Unlocking the opportunity for positive impact. *Nature Communications*, 11(1), 2468.
- Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine*, 25(1), 44-56.
- UNESCO. (2021). Recomendación sobre la Ética de la Inteligencia Artificial. Disponible en: <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics> (Acceso: 30-11-25)
- UNESCO. (2023). Readiness Assessment Methodology. Disponible en: <https://unesdoc.unesco.org/ark:/48223/pf0000385198> (Acceso: 30-11-25)
- UNESCO. (2024). Chile lanza política nacional de IA e introduce proyecto de ley siguiendo las recomendaciones de la UNESCO. Disponible en: <https://www.unesco.org/en/articles/chile-launches-national-ai-policy-and-introduces-ai-bill-following-unesco-recommendations-0> (Acceso: 30-11-25)

- Weerts, H., Dudík, M., Edgar, R., Jalali, A., Lutz, R., & Madaio, M. (2023). Fairlearn: Assessing and improving fairness of ai systems. *Journal of Machine Learning Research*, 24(257).
- White & Case LLP. (2024). Foster innovation or mitigate risk? AI regulation in Latin America. Disponible en: <https://www.whitecase.com/insight-our-thinking/latin-america-focus-2024-ai-regulation> (Acceso: 30-11-25)
- Winner, L. (1980). Do artifacts have politics?. *Daedalus*, 109(1), 121-136.
- Zhang, B., Anderljung, M., Kahn, L., Dreksler, N., Horowitz, M. C., & Dafoe, A. (2021). Ethics and governance of artificial intelligence: Evidence from a survey of machine learning researchers. *Journal of Artificial Intelligence Research*, 71, 591-666.
- Zuboff, S. (2023). The age of surveillance capitalism. In *Social theory re-wired* (pp. 203-213). Routledge.