Revista Hambatu Science

Ambato – Ecuador / ISSN 3151-815X (en línea) / julio - septiembre 2026

Volumen 1, Número 3

https://doi.org/10.63862/rhs-v1n3-275-294-2026

La inteligencia artificial como política social: regulación centrada en dignidad humana y formación de ciudadanía crítica

Artificial intelligence as social policy: regulation centered on human dignity and the formation of critical citizenship

César Eduardo Jiménez López
Universidad Autónoma de Guerrero

Maritza Leyssy Peña Gómez
Universidad Autónoma de Guerrero

Alberto Hernández Hernández
Instituto de Investigaciones Jurídicas, UNAM

Ricardo Alfonso Vásquez Alarcón
Escuela Judicial Electoral

Francisco Rosendo Olivares
Universidad Nacional Rosario Castellanos

Revista Hambatu Science
Periodicidad: julio - septiembre
Volumen: 1, Número: 3, Año: 2026

DOI: https://doi.org/10.63862/rhs-v1n3-275-294-2026

La inteligencia artificial como política social: regulación centrada en dignidad humana y formación de ciudadanía crítica

César Eduardo Jiménez López
Universidad Autónoma de Guerrero
09060888@uagro.mx
https://orcid.org/0009-0006-0366-4138
Acapulco, México

Ricardo Alfonso Vásquez Alarcón
Escuela Judicial Electoral
r.alfonsovasquez@gmail.com
https://orcid.org/0009-0006-7444-2903
Veracruz, México

Maritza Leyssy Peña Gómez
Universidad Autónoma de Guerrero
maritzamairani88@outlook.com
https://orcid.org/0009-0002-0732-4721
Chilpancingo, México

Francisco Rosendo Olivares
Universidad Nacional Rosario Castellanos
franciscorosendo31@rcastellanos.cdmx.gob.mx
https://orcid.org/0000-0002-5580-2401
Ciudad de México, México

Alberto Hernández Hernández
Instituto de Investigaciones Jurídicas, UNAM
hricardo24@hotmail.com
https://orcid.org/0009-0004-0477-0907
Ciudad de México

Recibido: 2026-07-15

Aceptado: 2026-07-17

Publicado: 2026-07-24

Resumen

El presente artículo analiza la inteligencia artificial como vector emergente de política social y examina las condiciones éticas, jurídicas y democráticas que deben orientar su regulación. A partir de una investigación cualitativa, teórico-documental y propositiva, se revisan fuentes doctrinales, normativas e institucionales sobre inteligencia artificial, derechos humanos, dignidad humana, política social, democracia deliberativa y ciudadanía crítica. El objetivo es identificar los riesgos de una incorporación tecnocrática de sistemas algorítmicos en la administración pública y proponer criterios mínimos para una regulación centrada en la persona. Los resultados muestran que la inteligencia artificial puede fortalecer la identificación de necesidades, la asignación de recursos y el seguimiento de programas sociales; sin embargo, también puede reproducir discriminación, vigilancia, opacidad, exclusión digital y debilitamiento de la deliberación pública. Como aportación, se formula una matriz de regulación basada en dignidad humana, transparencia, responsabilidad algorítmica, supervisión humana, igualdad sustantiva, protección de datos, alfabetización digital y participación ciudadana. Se concluye que la inteligencia artificial solo puede operar legítimamente como política social cuando amplía derechos, fortalece capacidades democráticas y permite que la ciudadanía comprenda, cuestione y participe en las decisiones tecnológicas que afectan su vida social.

Palabras clave: inteligencia artificial, política social, dignidad humana, derechos humanos, democracia deliberativa, ciudadanía crítica.

Revista Hambatu Science
Periodicidad: abril - junio
Volumen: 1, Número: 2, Año: 2026

Artificial intelligence as social policy: regulation centered on human dignity and the formation of critical citizenship

Abstract

This article analyzes artificial intelligence as an emerging vector of social policy and examines the ethical, legal, and democratic conditions that should guide its regulation. Based on a qualitative, theoretical-documentary, and propositional research design, the study reviews doctrinal, normative, and institutional sources on artificial intelligence, human rights, human dignity, social policy, deliberative democracy, and critical citizenship. Its objective is to identify the risks of a technocratic incorporation of algorithmic systems into public administration and to propose minimum criteria for person-centered regulation. The results show that artificial intelligence may strengthen the identification of needs, resource allocation, and monitoring of social programs; however, it may also reproduce discrimination, surveillance, opacity, digital exclusion, and the weakening of public deliberation. As a contribution, the article formulates a regulatory matrix based on human dignity, transparency, algorithmic accountability, human oversight, substantive equality, data protection, digital literacy, and citizen participation. It concludes that artificial intelligence can legitimately operate as social policy only when it expands rights, strengthens democratic capacities, and allows citizens to understand, question, and participate in technological decisions that affect their social life.

Keywords: artificial intelligence, social policy, human dignity, human rights, deliberative democracy, critical citizenship.

Introducción

La inteligencia artificial se ha incorporado de manera acelerada en múltiples dimensiones de la vida social. Su presencia ya no se limita al ámbito tecnológico o empresarial, sino que alcanza la administración pública, la educación, la salud, la seguridad, la justicia, la economía digital y el diseño de políticas sociales. Esta expansión obliga a comprenderla no únicamente como una herramienta técnica, sino como un fenómeno social, jurídico y político capaz de modificar la forma en que el Estado toma decisiones, distribuye recursos, identifica necesidades y organiza respuestas institucionales (Crawford, 2021; Floridi & Cowls, 2019).

En el campo público, los sistemas algorítmicos pueden utilizarse para automatizar trámites, analizar grandes volúmenes de información, detectar patrones de vulnerabilidad, focalizar programas, optimizar recursos y anticipar demandas sociales. Estos usos ofrecen ventajas administrativas relevantes, sobre todo en contextos donde las instituciones enfrentan saturación, dispersión territorial y déficit de información. Sin embargo, también plantean riesgos cuando las decisiones automatizadas operan sin controles, sin transparencia o sin mecanismos efectivos de revisión humana (Mittelstadt et al., 2016; O'Neil, 2016).

El problema central no radica en la existencia de la inteligencia artificial, sino en la forma en que se incorpora a las estructuras sociales e institucionales. Si la IA se adopta bajo una lógica exclusivamente eficientista, puede reducir los problemas públicos a datos procesables y convertir a las personas en objetos de clasificación, predicción o administración. En cambio, si se orienta desde la dignidad humana, los derechos humanos y la ciudadanía crítica, puede contribuir a políticas sociales más justas, inclusivas y democráticas (Nussbaum, 2012; Sen, 2000).

La dignidad humana constituye el criterio rector de esta discusión. Los derechos humanos descansan en la idea de que toda persona debe ser tratada como fin y no como medio, con independencia de su situación económica, territorial, cultural o digital. Por ello, cualquier política pública mediada por IA debe evaluarse no solo por su eficiencia, sino por su capacidad para proteger la igualdad, la autonomía, la privacidad, la no discriminación, el debido proceso y el acceso efectivo a derechos (Comisión Nacional de los Derechos Humanos [CNDH], s. f.; UNESCO, 2022).

En el ámbito internacional, diversos instrumentos han reconocido la necesidad de establecer marcos éticos y jurídicos para la inteligencia artificial. La UNESCO adoptó una recomendación que coloca la dignidad humana, los derechos humanos, la transparencia, la justicia, la supervisión humana y la responsabilidad como elementos centrales de la gobernanza tecnológica (UNESCO, 2022). La Organización para la Cooperación y el Desarrollo Económicos sostiene que los sistemas de IA deben ser confiables, centrados en las personas y respetuosos de los derechos humanos y los valores democráticos (Organización para la Cooperación y el Desarrollo Económicos [OCDE], 2024).

La Unión Europea aprobó el Reglamento (UE) 2024/1689, conocido como Reglamento de Inteligencia Artificial, que establece un marco jurídico basado en riesgos, con obligaciones diferenciadas según el impacto potencial de los sistemas sobre la seguridad, la salud y los derechos fundamentales (Unión Europea, 2024). A su vez, la resolución A/RES/78/265 de Naciones Unidas vincula los sistemas de IA seguros, protegidos y confiables con el desarrollo sostenible, la reducción de brechas digitales y la protección de derechos (Naciones Unidas, 2024). Estos referentes muestran que la IA ya no puede ser tratada solo como innovación, sino como objeto de regulación pública.

En México, el debate normativo sobre inteligencia artificial se encuentra en proceso de construcción. La propuesta de marco normativo discutida en el Senado de la República plantea una regulación flexible y adaptable basada en derechos humanos, supervisión, trazabilidad, gobernanza, combate a sesgos y responsabilidad en el uso de sistemas algorítmicos (Senado de la República, 2025). Este proceso constituye una oportunidad para formular un modelo propio sensible a las desigualdades sociales, territoriales y digitales del país.

Desde esta perspectiva, el presente artículo parte de la siguiente pregunta de investigación: ¿de qué manera la inteligencia artificial puede comprenderse como un vector de política social y qué elementos debe contener una regulación centrada en dignidad humana y formación de ciudadanía crítica? La hipótesis sostiene que la IA solo puede operar legítimamente como política social si se regula bajo principios de dignidad humana, transparencia, responsabilidad algorítmica, supervisión humana, igualdad sustantiva y deliberación pública.

El objetivo general consiste en analizar la inteligencia artificial como vector emergente de política social, a partir de la necesidad de construir una regulación centrada en la dignidad

humana y en la formación de ciudadanía crítica. Como objetivos específicos se plantean: identificar los principales riesgos de la IA frente a los derechos humanos; examinar los aportes de la democracia deliberativa para comprender la legitimidad de decisiones públicas mediadas por algoritmos; y proponer categorías mínimas para una regulación orientada a la justicia social.

Estado del arte

La literatura reciente sobre inteligencia artificial ha transitado de una visión predominantemente técnica hacia enfoques jurídicos, éticos, sociales y políticos. En sus primeras etapas, el debate se concentró en la automatización de tareas y en la capacidad de los sistemas computacionales para procesar información. Sin embargo, el desarrollo del aprendizaje automático, los sistemas predictivos y la IA generativa amplió la discusión hacia sus impactos en derechos humanos, democracia, trabajo, educación y políticas públicas (Crawford, 2021; Russell & Norvig, 2021).

En el ámbito de la política social, la IA puede entenderse como una herramienta que incide en la identificación de necesidades, la asignación de recursos, la focalización de beneficiarios, el diseño de programas y la evaluación de resultados. Una lectura optimista destaca su capacidad para mejorar eficiencia institucional, reducir tiempos administrativos y ampliar la capacidad del Estado para responder a problemas complejos. Una lectura crítica advierte que la automatización de decisiones públicas puede generar exclusiones injustificadas cuando los algoritmos operan con datos incompletos, sesgados o descontextualizados (Eubanks, 2018; O'Neil, 2016).

Los derechos humanos constituyen un eje central en esta discusión. La IA puede afectar privacidad, igualdad, no discriminación, acceso a servicios públicos, libertad de expresión, debido proceso y seguridad jurídica. En políticas sociales, estos riesgos adquieren especial gravedad porque las personas destinatarias de programas públicos suelen encontrarse en condiciones de vulnerabilidad. Una decisión automatizada errónea puede negar un apoyo, clasificar indebidamente a una persona, invisibilizar necesidades territoriales o reproducir criterios discriminatorios (Mittelstadt et al., 2016; UNESCO, 2022).

La UNESCO sostiene que la ética de la IA no debe limitarse a principios abstractos, sino traducirse en políticas concretas en materia de datos, género, educación, gobernanza, cooperación internacional y supervisión humana. Este punto es relevante porque permite pasar

de una ética declarativa a una ética institucionalizada. No basta afirmar que la IA debe ser justa o transparente; se requieren mecanismos verificables, autoridades competentes, responsabilidades definidas y vías de reparación frente a daños algorítmicos (Hernández Armenta & Rosendo Olivares, 2024).

La OCDE también establece principios orientados a una IA confiable, innovadora y respetuosa de derechos humanos y valores democráticos. Su enfoque resulta relevante porque vincula innovación con rendición de cuentas, robustez, transparencia y centralidad de la persona (OCDE, 2024). En consecuencia, la gobernanza de la IA no debe entenderse como una carga externa al desarrollo tecnológico, sino como condición para que la innovación sea socialmente legítima y jurídicamente responsable.

La Asamblea General de las Naciones Unidas adoptó la resolución A/RES/78/265 sobre sistemas de IA seguros, protegidos y confiables para el desarrollo sostenible. En ella se reconoce la necesidad de cerrar brechas digitales, fortalecer capacidades, respetar derechos humanos y orientar la IA hacia los Objetivos de Desarrollo Sostenible (Naciones Unidas, 2024). Esta resolución permite interpretar la IA como asunto global vinculado a justicia social, desarrollo, educación, cooperación y equidad tecnológica.

En el plano regulatorio, el Reglamento de Inteligencia Artificial de la Unión Europea representa una referencia importante por su enfoque basado en riesgos. Este modelo distingue entre prácticas prohibidas, sistemas de alto riesgo, obligaciones de transparencia y reglas aplicables a modelos de propósito general (Unión Europea, 2024). Para el análisis de políticas sociales, dicho enfoque resulta útil porque ciertos usos de IA en salud, educación, empleo, seguridad social o acceso a servicios públicos pueden producir impactos directos sobre derechos fundamentales.

No obstante, la regulación de la IA no debe limitarse a importar modelos extranjeros. En contextos latinoamericanos, la desigualdad social, la informalidad laboral, las brechas digitales, la violencia estructural y las diferencias territoriales obligan a formular una regulación situada (Ortega Cruz et al., 2026). Una política algorítmica aparentemente neutral puede generar efectos diferenciados sobre comunidades indígenas, personas con discapacidad, mujeres, niñas, niños, adolescentes, personas migrantes o sectores empobrecidos. Por ello, la regulación debe dialogar con el principio de igualdad sustantiva (Ferrajoli, 2011; Nussbaum, 2012).

Desde la teoría social, Jürgen Habermas ofrece una herramienta especialmente útil. Su teoría de la acción comunicativa plantea que la racionalidad social no debe reducirse a la racionalidad instrumental, es decir, a la elección de medios eficaces para alcanzar fines previamente definidos. Frente a ello, propone una racionalidad comunicativa basada en entendimiento, argumentación y posibilidad de participación en procesos de formación de voluntad colectiva (Habermas, 1987).

En Facticidad y validez, Habermas sostiene que la legitimidad del derecho democrático depende de procedimientos discursivos en los que las personas puedan reconocerse como destinatarias y autoras de las normas (Habermas, 1998). Aplicado al campo de la IA, esto implica que las reglas sobre automatización, datos, vigilancia, sistemas predictivos o decisiones algorítmicas no deben diseñarse únicamente por especialistas técnicos, empresas o autoridades administrativas, sino abrirse a deliberación pública.

El estado del arte permite identificar un vacío relevante: gran parte de la discusión sobre IA se concentra en innovación, eficiencia y productividad, mientras todavía resulta necesario fortalecer enfoques que la analicen como política social desde la dignidad humana, la justicia social y la ciudadanía crítica. Este artículo se ubica en ese espacio al proponer que la IA debe comprenderse como instrumento de intervención pública que requiere legitimidad democrática, regulación jurídica y apropiación ciudadana.

Metodología

La investigación adoptó un enfoque cualitativo de tipo teórico-documental, con alcance analítico y propositivo. Este enfoque resulta pertinente porque el objetivo no consiste en medir estadísticamente el impacto de la inteligencia artificial, sino en analizar críticamente sus implicaciones jurídicas, sociales y políticas a partir de fuentes doctrinales, normativas e institucionales. El estudio busca construir una interpretación argumentativa sobre la IA como vector de política social y sobre la necesidad de una regulación centrada en la dignidad humana (Hernández Sampieri & Mendoza Torres, 2018).

El método empleado fue analítico-crítico. Fue analítico porque descompone el objeto de estudio en categorías específicas: inteligencia artificial, política social, dignidad humana, derechos humanos, regulación, democracia deliberativa y ciudadanía crítica. Fue crítico porque no se

limita a describir beneficios de la IA, sino que examinó riesgos, tensiones y efectos adversos en contextos de desigualdad social, brecha digital y debilidad normativa.

La técnica principal fue el análisis documental. Se revisaron fuentes doctrinales relacionadas con democracia deliberativa, acción comunicativa, derechos humanos, política social y ética algorítmica. Asimismo, se incorporaron investigaciones recientes sobre ética de la IA en el ámbito jurídico, sesgos algorítmicos y justicia social en América Latina, así como control de convencionalidad y restricciones de derechos en decisiones mediadas por inteligencia artificial (Hernández Armenta & Rosendo Olivares, 2024; Ortega Cruz et al., 2026; Padilla Sanabria & Rosendo Olivares, 2026). También se consideraron fuentes institucionales internacionales, como la Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO, los Principios de IA de la OCDE, la resolución de Naciones Unidas sobre IA para el desarrollo sostenible, el Reglamento Europeo de Inteligencia Artificial y la discusión mexicana sobre regulación algorítmica (Naciones Unidas, 2024; OCDE, 2024; Senado de la República, 2025; UNESCO, 2022; Unión Europea, 2024).

El corpus documental se seleccionó a partir de cuatro criterios. Primero, relevancia temática, considerando documentos directamente relacionados con IA, derechos humanos, regulación y política social. Segundo, autoridad institucional, priorizando fuentes emitidas por organismos internacionales, órganos legislativos, instituciones académicas y editoriales especializadas. Tercero, pertinencia teórica, incorporando autores que fundamentan la relación entre legitimidad democrática, deliberación pública y ciudadanía crítica. Cuarto, actualidad, tomando en cuenta que el campo de la IA se transforma de manera constante.

El procedimiento de análisis se desarrolló en cuatro etapas. En la primera, se identificaron los usos de IA con incidencia en política social: focalización, evaluación de necesidades, administración de datos, prestación de servicios, vigilancia, predicción y automatización de decisiones. En la segunda, se clasificaron los riesgos en dimensiones jurídicas, sociales, democráticas y educativas. En la tercera, se contrastaron esos riesgos con estándares de dignidad humana, derechos humanos, gobernanza y deliberación. En la cuarta, se construyeron matrices de resultados y criterios regulatorios.

Debido a su naturaleza documental, la investigación no utilizó encuestas, entrevistas ni instrumentos estadísticos. Su principal limitación consiste en que no mide empíricamente la

percepción ciudadana ni el funcionamiento de un sistema específico de IA. No obstante, su fortaleza radica en articular teoría social, estándares internacionales y análisis jurídico para formular criterios transferibles a futuras investigaciones empíricas, lineamientos institucionales o propuestas normativas.

Tabla 1.
Categorías de análisis utilizadas en la investigación

Categoría de análisis	Finalidad analítica
Inteligencia artificial	Comprender su función como herramienta técnica y como fenómeno sociopolítico con efectos institucionales.
Política social	Examinar su impacto en la distribución de derechos, oportunidades, recursos y servicios públicos.
Dignidad humana	Establecer el criterio rector para evaluar límites, finalidades y efectos de la IA.
Derechos humanos	Identificar riesgos frente a igualdad, privacidad, no discriminación, debido proceso y acceso a derechos.
Democracia deliberativa	Analizar la legitimidad de decisiones públicas mediadas por sistemas algorítmicos.
Ciudadanía crítica	Valorar la necesidad de alfabetización digital, participación, comprensión y control democrático.

Fuente: *elaboración propia*.

Resultados

Los resultados del análisis documental muestran, en primer lugar, que la inteligencia artificial no puede comprenderse únicamente como herramienta auxiliar de gestión pública. Su capacidad para clasificar información, predecir comportamientos, automatizar decisiones y orientar intervenciones estatales la convierte en un factor que puede modificar la forma en que se diseñan y ejecutan políticas sociales. Esta conclusión coincide con los debates contemporáneos sobre el poder de los sistemas algorítmicos para ordenar, jerarquizar y administrar poblaciones (Crawford, 2021; Eubanks, 2018).

El primer hallazgo consiste en reconocer que la IA posee una dimensión social que rebasa su carácter técnico. Aunque su funcionamiento se basa en modelos matemáticos, bases de datos, algoritmos y sistemas computacionales, sus efectos se proyectan sobre personas concretas. Cuando un sistema automatizado interviene en educación, salud, empleo, seguridad o

programas sociales, no solo procesa información: también produce consecuencias sobre trayectorias de vida.

El segundo hallazgo muestra que la IA puede convertirse en mecanismo de inclusión o exclusión. La IA puede facilitar el acceso a servicios públicos o, por el contrario, negar beneficios cuando opera con datos insuficientes o criterios opacos; mejora la detección de necesidades sociales al tiempo que puede invisibilizar a quienes no figuran en bases de datos oficiales; y, si las autoridades trasladan sus decisions a sistemas automatizados, fortalice capacidades institucionales a costa de diluir la responsabilidad humana (O'Neil, 2016).

El tercer hallazgo se relaciona con los riesgos frente a los derechos humanos. La automatización puede afectar la dignidad humana cuando reduce a las personas a perfiles, puntajes o probabilidades. Este riesgo se intensifica en contextos de vulnerabilidad, donde las personas dependen de apoyos estatales, servicios educativos, programas de salud o mecanismos de protección social. Por ello, la IA debe evaluarse desde su impacto sobre derechos y no únicamente desde su utilidad administrativa (CNDH, s. f.; UNESCO, 2022).

Entre los riesgos más relevantes se identifican discriminación algorítmica, opacidad decisional, vigilancia excesiva, uso indebido de datos personales, exclusión digital, afectación al debido proceso y debilitamiento de la autonomía personal. La OCDE sostiene que los sistemas de IA deben respetar derechos humanos, valores democráticos y principios centrados en las personas durante todo su ciclo de vida, incluyendo igualdad, privacidad, dignidad, autonomía y justicia social (OCDE, 2024).

El cuarto hallazgo permite advertir que una regulación centrada solo en aspectos técnicos resulta insuficiente. Aunque son importantes la seguridad, la calidad de datos, la interoperabilidad y la evaluación de riesgos, estos elementos no bastan para garantizar justicia social. La regulación debe preguntar para qué se usa el sistema, a quién afecta, qué derechos puede comprometer y bajo qué procedimiento puede cuestionarse una decisión automatizada (Floridi & Cowls, 2019; Mittelstadt et al., 2016).

Cuando una decisión algorítmica estatal limita, condiciona o afecta el ejercicio de un derecho, su evaluación no debería agotarse en una revisión técnica del sistema. También debe determinarse si la medida cuenta con fundamento jurídico, persigue una finalidad legítima,

resulta idónea y necesaria, y mantiene una relación proporcional con el derecho afectado. Desde esta perspectiva, los parámetros del test de restricción desarrollados en el Sistema Interamericano pueden servir como herramienta de control de convencionalidad para examinar decisiones públicas mediadas por inteligencia artificial, especialmente cuando inciden en el debido proceso, la igualdad o el acceso efectivo a derechos (Padilla Sanabria & Rosendo Olivares, 2026).

El quinto hallazgo destaca la importancia de la supervisión humana. La IA no debe sustituir la responsabilidad de las autoridades públicas. En materia de política social, toda decisión que afecte derechos debe conservar una instancia humana capaz de revisar, corregir, explicar y justificar el resultado. La supervisión humana no debe ser simbólica, sino efectiva, documentada y jurídicamente vinculante (UNESCO, 2022; Unión Europea, 2024).

El sexto hallazgo sostiene que la alfabetización digital y la formación de ciudadanía crítica son condiciones necesarias para la gobernanza democrática de la IA. Si la ciudadanía desconoce cómo operan los sistemas algorítmicos, difícilmente podrá cuestionarlos, exigir transparencia o participar en la construcción de marcos regulatorios. La brecha digital no solo implica falta de acceso a dispositivos o internet; también supone desigualdad en la capacidad de comprender, evaluar y disputar el uso público de tecnologías inteligentes (Naciones Unidas, 2024; Zuboff, 2019).

Tabla 2.
Riesgos de la inteligencia artificial como política social y garantías regulatorias mínimas

Riesgo identificado	Efecto en política social	Garantía regulatoria mínima
Discriminación algorítmica	Exclusión injustificada de personas o grupos en el acceso a programas, servicios o beneficios públicos.	Auditoría de sesgos, enfoque de igualdad sustantiva, pruebas con grupos diversos y ajustes razonables.
Opacidad decisional	Imposibilidad de conocer por qué una persona fue clasificada, excluida, priorizada o vigilada.	Transparencia, explicabilidad, trazabilidad y derecho a recibir razones comprensibles.
Vigilancia excesiva	Normalización del control social y afectación de privacidad, autonomía y libertad.	Proporcionalidad, finalidad legítima, minimización de datos y límites claros de monitoreo.
Desplazamiento de responsabilidad	Autoridades que atribuyen la decisión al sistema y evaden justificación pública.	Responsabilidad institucional, supervisión humana efectiva e identificación de personas responsables.
Brecha digital y ciudadanía débil	Personas sin capacidades para comprender, cuestionar o impugnar decisiones automatizadas.	Alfabetización digital, educación algorítmica y mecanismos accesibles de participación.
Tecnocratización de la política social	Reducción de problemas estructurales a datos procesables y soluciones instrumentales.	Deliberación pública, evaluación social de impacto y participación de comunidades afectadas.

Fuente: *elaboración propia con base en Eubanks (2018), O'Neil (2016), UNESCO (2022), OCDE (2024), Naciones Unidas (2024) y Unión Europea (2024).*

A partir de estos hallazgos, se propone una matriz regulatoria que permite trasladar los principios éticos a criterios operativos. La regulación no debe limitarse a enunciar valores, sino establecer obligaciones verificables en dignidad humana, transparencia, responsabilidad algorítmica, supervisión humana, igualdad sustantiva, protección de datos, alfabetización digital y deliberación pública para instituciones que diseñen, adquieran o utilicen sistemas de IA en contextos de política social.

Tabla 3.
Matriz de regulación centrada en dignidad humana y ciudadanía crítica

Principio	Pregunta de control	Criterio de implementación
Dignidad humana	¿El sistema trata a la persona como sujeto de derechos y no solo como dato o perfil?	Prohibir usos que cosifiquen, discriminen o reduzcan derechos a puntuaciones automatizadas.
Transparencia y explicabilidad	¿La ciudadanía puede comprender la finalidad, datos y criterios generales del sistema?	Publicar información accesible, lenguaje claro, documentación y vías de explicación individual.
Responsabilidad algorítmica	¿Existe una autoridad o sujeto responsable por el diseño, adquisición, operación y daño?	Definir obligaciones, auditorías, bitácoras, responsabilidades administrativas y reparación.
Supervisión humana efectiva	¿Una persona puede revisar y modificar decisiones automatizadas relevantes?	Garantizar revisión humana previa o posterior en decisiones que afecten derechos.
Igualdad sustantiva	¿El sistema produce impactos diferenciados sobre grupos históricamente excluidos?	Evaluar sesgos, accesibilidad, brechas digitales y efectos territoriales antes del despliegue.
Alfabetización digital crítica	¿La población cuenta con capacidades para entender y cuestionar la IA pública?	Incluir educación algorítmica, formación ética y campañas ciudadanas de derechos digitales.
Deliberación pública	¿Las comunidades afectadas participaron en la definición de usos, límites y controles?	Crear consultas, mecanismos participativos, comités plurales y revisión democrática.
Protección de datos personales	¿La recolección, tratamiento y conservación de datos respetan finalidad legítima y minimización?	Limitar datos necesarios, establecer seguridad, confidencialidad, plazos de conservación y vías de rectificación o eliminación.

Fuente: *elaboración propia con base en Habermas (1987, 1998), UNESCO (2022), OCDE (2024), Naciones Unidas (2024), Mittelstadt et al. (2016), Zuboff (2019) y Senado de la República (2025).*

Discusión

Los resultados permiten sostener que la inteligencia artificial ocupa una posición ambivalente en las sociedades contemporáneas. Por un lado, representa una herramienta con capacidad para mejorar procesos institucionales, fortalecer diagnósticos y ampliar la respuesta estatal. Por otro lado, puede convertirse en mecanismo de reproducción de desigualdades si se implementa sin regulación, sin deliberación y sin enfoque de derechos humanos (Crawford, 2021; Eubanks, 2018).

Desde la teoría de Habermas, esta tensión puede comprenderse como confrontación entre racionalidad instrumental y racionalidad comunicativa. La racionalidad instrumental busca

eficacia, cálculo, predicción y control. Es necesaria en ámbitos técnicos, pero resulta insuficiente para legitimar decisiones públicas. La racionalidad comunicativa exige argumentación, participación, entendimiento y posibilidad de crítica (Habermas, 1987).

Cuando la IA se introduce en política social desde una lógica puramente instrumental, el Estado puede considerar que una decisión es válida porque el sistema la produjo de manera rápida o porque se basa en datos. Sin embargo, desde una perspectiva deliberativa, la legitimidad no depende solo de eficiencia, sino de la posibilidad de justificar públicamente las decisiones, escuchar a las personas afectadas y someter los criterios técnicos a control democrático (Habermas, 1998).

La dignidad humana aparece entonces como límite y finalidad. Como límite, impide que las personas sean tratadas únicamente como datos, perfiles o medios para optimizar procesos administrativos. Como finalidad, orienta el desarrollo tecnológico hacia la ampliación de capacidades, derechos y condiciones de vida. Una IA compatible con la dignidad humana no debe sustituir el juicio ético ni la responsabilidad pública, sino fortalecerlos (CNDH, s. f.; Nussbaum, 2012).

La política social no puede reducirse a distribución automática de beneficios. Su finalidad consiste en atender desigualdades, garantizar derechos, fortalecer cohesión social y construir condiciones para una vida digna. Por ello, la incorporación de IA en este campo debe analizarse con especial cuidado. Un error algorítmico en una plataforma comercial puede afectar una preferencia de consumo; un error algorítmico en política social puede negar alimentación, salud, educación, vivienda o protección a una persona.

La discusión jurídica refuerza la necesidad de una regulación basada en riesgo, pero también situada. El modelo europeo muestra la utilidad de clasificar usos según su impacto potencial (Unión Europea, 2024). Sin embargo, en América Latina resulta indispensable añadir criterios de desigualdad estructural, brecha digital, diversidad cultural y vulnerabilidad territorial. De lo contrario, una regulación formalmente neutral puede convertirse en una nueva fuente de exclusión.

La ciudadanía crítica no es un elemento accesorio, sino una condición estructural de la regulación. Si las personas no comprenden los efectos sociales de la IA, se produce una

asimetría entre quienes diseñan sistemas y quienes los padecen. Esa asimetría debilita la democracia porque desplaza decisiones públicas hacia espacios técnicos poco accesibles (Zuboff, 2019).

En ese sentido, la alfabetización digital debe entenderse como política social. No se trata únicamente de enseñar a usar herramientas tecnológicas, sino de formar capacidades para comprender riesgos, exigir derechos, identificar sesgos, proteger datos personales y participar en debates públicos sobre el uso de IA. La ciudadanía crítica implica que las personas no sean usuarias pasivas de tecnología, sino sujetos capaces de intervenir en su orientación normativa (Naciones Unidas, 2024; UNESCO, 2022).

La discusión mexicana confirma la oportunidad de construir una regulación propia. El debate legislativo sobre IA no debería limitarse a promover innovación o competitividad, sino atender las condiciones sociales del país: brecha de conectividad, desigualdad territorial, desconfianza institucional, discriminación histórica y capacidades administrativas desiguales (Senado de la República, 2025). Esto exige un modelo regulatorio gradual, flexible, democrático y centrado en la persona.

Una limitación del presente estudio es su carácter documental. No se analizan sistemas específicos ni se mide la percepción de comunidades afectadas. Esta limitación abre futuras líneas de investigación: estudios de caso sobre IA en programas sociales, análisis de impacto en poblaciones vulnerables, evaluación de marcos regulatorios locales y diseño de indicadores de alfabetización algorítmica. También resulta necesario estudiar cómo integrar mecanismos de participación ciudadana en la adquisición y uso de sistemas públicos de IA.

Conclusiones

La inteligencia artificial constituye uno de los fenómenos tecnológicos más relevantes de la vida contemporánea. Su impacto no se limita al ámbito técnico, sino que alcanza dimensiones jurídicas, sociales, políticas y éticas. En el campo de la política social puede contribuir a mejorar diagnósticos, optimizar recursos, ampliar capacidades institucionales y fortalecer la atención a problemas públicos complejos. Sin embargo, también puede generar riesgos significativos cuando opera sin regulación clara, sin supervisión humana y sin criterios de dignidad humana.

El análisis permite concluir que la IA debe comprenderse como vector de política social. Esto

significa que no solo sirve para automatizar tareas, sino que puede incidir en la forma en que el Estado distribuye oportunidades, reconoce necesidades y garantiza derechos. Por ello, su incorporación en la administración pública y en programas sociales debe sujetarse a controles democráticos, jurídicos y éticos (Ferrajoli, 2011; Sen, 2000).

La dignidad humana debe ocupar el centro de la regulación. Una política pública mediada por IA no puede tratar a las personas como simples datos o perfiles de riesgo. Todo sistema algorítmico que afecte derechos debe respetar igualdad, no discriminación, privacidad, autonomía, debido proceso y posibilidad de revisión humana. La dignidad humana no es un elemento decorativo del discurso jurídico, sino el parámetro que permite evaluar la legitimidad del desarrollo tecnológico (CNDH, s. f.; UNESCO, 2022).

Desde la perspectiva de Habermas, la regulación de la IA exige deliberación pública. Las decisiones sobre tecnologías que afectan la vida social no deben quedar únicamente en manos de especialistas, empresas o autoridades administrativas. La ciudadanía debe participar en la definición de límites, finalidades, usos permitidos, mecanismos de control y vías de reparación. Sin participación ciudadana, la regulación corre el riesgo de convertirse en una arquitectura técnica sin legitimidad democrática (Habermas, 1998).

La alfabetización digital y algorítmica debe asumirse como parte de la política social. Una ciudadanía crítica requiere herramientas para comprender cómo operan los sistemas de IA, qué riesgos producen y qué derechos pueden exigirse frente a decisiones automatizadas. La educación tecnológica no debe limitarse al uso instrumental de plataformas, sino incluir formación ética, jurídica y democrática (Naciones Unidas, 2024; OCDE, 2024).

En el caso mexicano, la discusión sobre IA representa una oportunidad para construir un marco normativo propio, sensible a las desigualdades del país y orientado a la justicia social. No basta importar modelos regulatorios extranjeros. Es necesario diseñar una regulación que considere brechas digitales, diversidad cultural, derechos humanos, capacidades institucionales y participación ciudadana (Senado de la República, 2025).

Finalmente, la investigación confirma que la IA puede contribuir a la construcción de una ciudadanía informada, crítica y corresponsable siempre que su desarrollo se oriente hacia fines democráticos. Regular la inteligencia artificial no significa frenar la innovación, sino darle

dirección ética y social. La verdadera pregunta no es si debe utilizarse IA, sino bajo qué principios, con qué controles, para quiénes y con qué finalidad. Una IA centrada en la dignidad humana puede fortalecer la política social; una IA sin deliberación puede debilitar la democracia.

Referencias bibliográficas

Comisión Nacional de los Derechos Humanos. (s. f.). ¿Qué son los derechos humanos? https://www.cndh.org.mx/derechos-humanos/que-son-los-derechos-humanos/

Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.

Ferrajoli, L. (2011). Principia iuris: Teoría del derecho y de la democracia. Trotta.

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

Habermas, J. (1987). Teoría de la acción comunicativa (Vols. 1-2). Taurus.

Habermas, J. (1998). Facticidad y validez: Sobre el derecho y el Estado democrático de derecho en términos de teoría del discurso. Trotta.

Hernández Armenta, C., & Rosendo Olivares, F. (2024). Aspectos éticos en la aplicación de inteligencia artificial para la investigación jurídica. Pensamiento Crítico. Revista de Investigación Multidisciplinaria, 10(19), 1-14. https://doi.org/10.64040/nyyryn02

Hernández Sampieri, R., & Mendoza Torres, C. P. (2018). Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta. McGraw-Hill Education.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 1-21. https://doi.org/10.1177/2053951716679679

Naciones Unidas. (2024). Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development (A/RES/78/265). Asamblea General de las Naciones Unidas. https://digitallibrary.un.org/record/4043244

Nussbaum, M. C. (2012). Crear capacidades: Propuesta para el desarrollo humano. Paidós.

O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown.

Organización para la Cooperación y el Desarrollo Económicos. (2024). OECD AI principles. OECD. https://www.oecd.org/en/topics/sub-issues/ai-principles.html

Ortega Cruz, Y. del S., Rosendo Olivares, F., & Portillo Juárez, J. T. (2026). Justicia social frente a la IA: sesgos y vulneraciones de derechos humanos en México, Brasil y Colombia. Tramas y Redes, (10), 163-180. https://doi.org/10.54871/cl4c10ai

Padilla Sanabria, L., & Rosendo Olivares, F. (2026). El test de restricción y la inteligencia artificial: Un análisis convencional de los derechos humanos en los procesos penales a partir del Sistema Interamericano de Derechos Humanos. Revista del Posgrado en Derecho de la UNAM, (e1), 163. https://doi.org/10.22201/ppd.26831783e.2026.e1.482

Russell, S., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.

Sen, A. (2000). Desarrollo y libertad. Planeta.

Senado de la República. (2025). Propuesta de marco normativo para la inteligencia artificial (IA) en México. Comisión de Análisis, Seguimiento y Evaluación sobre la Aplicación y Desarrollo de la Inteligencia Artificial en México. https://comisiones.senado.gob.mx/inteligencia_artificial/images/noticias/Propuesta.pdf

UNESCO. (2022). Recomendación sobre la ética de la inteligencia artificial. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. https://unesdoc.unesco.org/ark:/48223/pf0000381137

Unión Europea. (2024). Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo de 13 de junio de 2024 por el que se establecen normas armonizadas en materia de

inteligencia artificial. Diario Oficial de la Unión Europea. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.

Declaraciones finales

Conflicto de intereses: Los autores declaran que no existe conflicto de interés posible.

Financiamiento: No existió asistencia financiera de partes externas al presente artículo.

Agradecimiento: N/A

Nota editorial: El artículo no deriva de una publicación anterior ni se encuentra sometido simultáneamente a otra revista.