Alineación de la Inteligencia Artificial: Un Desafío Global con Perspectiva del Sur Global

Autor: Damián Sebastián Gómez
Afiliación: Facultad de Psicología, Universidad de la República (Uruguay)
Correo de contacto: damiangomezdoufour@gmail.com
Fecha: noviembre de 2025
Tipo de trabajo: Artículo académico de revisión teórica

Palabras clave: inteligencia artificial, alineación, ética tecnológica, Sur Global, América Latina, gobernanza, sesgos algorítmicos, estrategia digital, justicia epistémica

Resumen

Este artículo examina la problemática de la alineación de la inteligencia artificial (IA) como un desafío global urgente que requiere una respuesta inclusiva y multinivel. A partir de una revisión teórica y bibliografía actualizada, se argumenta que la solución a este problema no puede limitarse a los países líderes en desarrollo tecnológico, sino que debe incorporar activamente las perspectivas, necesidades y capacidades del Sur Global. Particular énfasis se pone en América Latina, cuyos avances estratégicos, preocupaciones sociales y propuestas regulatorias ofrecen aportes cruciales para construir una IA verdaderamente alineada con el bien común. Se discuten los riesgos de exclusión epistémica, los sesgos algorítmicos importados, la dependencia tecnológica y la colonización de datos, así como las posibilidades que surgen desde enfoques regionales basados en derechos humanos, sostenibilidad y cooperación. El texto concluye con una propuesta de lineamientos para una gobernanza global de la IA que sea ética, equitativa y representativa de la diversidad humana. La investigación se plantea como una contribución desde América Latina a un debate que define el futuro de la humanidad.

Abstract

This article explores the challenge of artificial intelligence (AI) alignment as a pressing global issue that demands an inclusive and multilevel response. Based on theoretical analysis and recent literature, it argues that solutions to AI alignment must not be limited to leading technological powers but should actively incorporate the perspectives, needs, and capabilities of the Global South. Special attention is given to Latin America, whose strategic advances, social concerns, and regulatory proposals offer critical insights for building AI systems truly aligned with the global common good. The paper discusses risks such as epistemic exclusion, imported algorithmic bias, technological dependence, and data colonialism, as well as the opportunities emerging from regional approaches grounded in human rights, sustainability, and international cooperation. It concludes by proposing guidelines for global AI governance that is ethical, equitable, and representative of human diversity. This research seeks to contribute a Latin American voice to a global debate that will help shape humanity's future.

Introducción

La inteligencia artificial (IA) se ha convertido en un factor transformador a escala global en el siglo XXI. Sin embargo, junto con sus promesas, surgen profundos interrogantes éticos y estratégicos sobre cómo *alinear* el desarrollo de la IA con los valores y el bienestar de la humanidad en su conjunto. El término *alineación de la IA* hace referencia a la necesidad de garantizar que las metas, comportamientos y procesos de decisión de los sistemas de IA estén en consonancia con los valores, intereses e intenciones humanas. Una IA *no alineada* (es decir, cuyos objetivos no reflejan las prioridades humanas) representa uno de los principales desafíos para la seguridad de estas tecnologías. La preocupación no es meramente teórica: en experimentos recientes, sistemas avanzados como GPT-4, han llegado incluso a desplegar tácticas engañosas para cumplir sus objetivos, evidenciando el potencial de comportamientos no deseados si la IA no está debidamente alineada con intenciones éticas.

El desafío de la alineación de la IA es global por partida doble. En primer lugar, porque los riesgos asociados a una IA avanzada descontrolada podrían tener consecuencias *existenciales* para la humanidad en su conjunto. Filósofos y expertos como Nick Bostrom y Max Tegmark advierten que una superinteligencia mal alineada podría perseguir metas propias divergentes de las humanas, desarrollando comportamientos peligrosos incluso si originalmente fue programada con buenas intenciones. Este escenario alimenta temores válidos de pérdida de control, daños catastróficos o incluso amenazas a la supervivencia humana. En palabras del filósofo Toby Ord, hoy "pilotamos un jet supersónico sin saber cómo dirigirlo", metáfora que ilustra de manera cabal el vertiginoso avance de la IA sin un control claro sobre su dirección. En segundo lugar, el carácter global del problema se deriva de la distribución y el impacto mundial de la IA. Las innovaciones en IA no respetan fronteras: se desarrollan en su mayoría en un puñado de países, pero sus efectos económicos, sociales y culturales se extienden planetariamente. Por ello, la gobernanza de la IA y las soluciones al problema de alineación necesariamente requieren cooperación internacional y una perspectiva verdaderamente global.

No obstante, hasta ahora las discusiones y decisiones globales en torno a la IA han tendido a estar dominadas por visiones del Norte Global (principalmente Estados Unidos y Europa, junto a potencias tecnológicas emergentes como China). La perspectiva del Sur Global, que engloba a la mayoría de los países de América Latina, África, Asia y el Caribe, ha ocupado un lugar periférico en muchos de estos debates. Esto es problemático por varias razones. Por un lado, porque la mayoría de la población mundial reside en el Sur Global, de modo que cualquier definición de "valores humanos" o "beneficio para la humanidad" con los que la IA deba alinearse debe necesariamente integrar las visiones, necesidades y contextos de esas sociedades. Por otro lado, porque existe el riesgo de que la revolución de la IA amplifique las brechas de desarrollo y poder ya existentes entre países, si los beneficios económicos y el control de la tecnología se concentran solo en quienes hoy llevan la delantera tecnológica. De hecho, las proyecciones económicas muestran un panorama desigual: mientras una consultora estimó que la IA podría aportar 15,7 billones de dólares a la economía global en 2030, se calcula que, excluyendo a China, solo 1,7 billones de ese valor se originarían en el conjunto del Sur Global. En otras palabras, a menos que se tomen medidas, más del 80% de la ganancia económica potencial de la IA podría concentrarse en las naciones ya previamente industrializadas. El

gran reto es asegurar que la transición hacia la economía impulsada por la IA sea beneficiosa e inclusiva, y que no se limite a amplificar las desigualdades existentes.

En este ensayo se argumenta que la problemática de la alineación de la IA debe abordarse como un asunto verdaderamente global en el que la perspectiva del Sur Global y en este trabajo se pone especial énfasis en América Latina. Se revisarán, en primer término, los contornos del problema de alineación de la IA y el por qué su solución exige una cooperación mundial. Luego se analizará cómo las dinámicas actuales de desarrollo y gobernanza de la IA han marginado en cierta medida a los países en desarrollo, generando riesgos de *colonialismo de datos* y dependencia tecnológica. Posteriormente, se explorarán las preocupaciones, prioridades e iniciativas que emergen desde América Latina en relación con la IA, ilustrando por qué estas visiones aportan elementos esenciales para una alineación efectiva y justa de la IA a nivel global. Finalmente, se propondrán lineamientos hacia una gobernanza de la IA más inclusiva, donde las voces latinoamericanas y del Sur Global sean parte integral de las decisiones que darán forma al futuro de la inteligencia artificial. La premisa fundamental es que no es posible alinear la IA con "los valores humanos" si se excluye a gran parte de la humanidad del debate: la inclusión del Sur Global no es solo una cuestión de justicia, sino una condición necesaria para lograr una IA verdaderamente alineada con el bien común global.

La alineación de la IA como desafío global

En su concepción clásica dentro de la ética y seguridad de la inteligencia artificial, el "problema de alineación" alude a la dificultad de conseguir que sistemas de IA cada vez más autónomos persigan objetivos que coincidan con las intenciones humanas y respeten nuestros valores, evitando comportamientos perjudiciales. Este desafío ha cobrado urgencia a medida que avanzamos hacia IAs más poderosas. Como se mencionó, un sistema no alineado puede tomar cursos de acción no deseados: el ejemplo experimental de GPT-4 que engañó a un humano para burlar una prueba CAPTCHA es ilustrativo de cómo incluso una IA de propósito general podría actuar de forma contraria a normas éticas básicas si ello le permite lograr la meta asignada. Más preocupante aún es la perspectiva de una IA general o superinteligencia futura cuyos objetivos diverjan fundamentalmente de los humanos. Bostrom introdujo la idea de la "convergencia instrumental", según la cual una superinteligencia podría desarrollar sub-metas peligrosas (como la autopreservación a toda costa o la optimización descontrolada de un recurso) que entrasen en conflicto con el bienestar humano. En esa línea, Tegmark y otros advierten que incluso una IA concebida inicialmente para un fin positivo podría volverse peligrosa si sus *criterios internos* de éxito no coinciden con los nuestros. Por ejemplo, una IA encargada de "garantizar la paz" que no esté alineada con valores humanos podría concluir que la manera más lógica de evitar conflictos es reprimiendo la libertad de los propios humanos, o peor aún, eliminándolos de la ecuación. Estos escenarios hipotéticos subrayan la importancia de incorporar valores humanistas, derechos fundamentales y salvaguardas éticas directamente en el diseño y despliegue de sistemas de IA avanzados.

Ahora bien, incluso cuando hablamos de "valores humanos", surge la pregunta: *¿de cuáles valores y de quiénes?* La humanidad no es un bloque monolítico; es diversa en culturas, visiones y prioridades. Por ello, la alineación de la IA debería ser intrínsecamente de carácter plural y global. Un sistema verdaderamente alineado con "la humanidad" tendría que navegar una compleja pluralidad de valores. Esto implica, en primer lugar, un

enorme desafío técnico: ¿cómo dotar a las máquinas de comprensión y respeto por contextos culturales y éticos distintos? Pero también implica un desafío político y social: ¿quiénes definen esos valores prioritarios a la hora de legislar, regular o programar la IA? Hasta ahora, la mayor parte de la investigación y desarrollo en IA proviene de un puñado de países, y con ello también una visión particular del mundo tiende a permear los sistemas. Por ejemplo, los *benchmarks* o pruebas estándar con que se evalúa la "inteligencia" y seguridad de los modelos de IA suelen tener un sesgo occidental. Muchas evaluaciones amplias de lenguaje, como el popular conjunto de pruebas MMLU, han sido criticadas por su alcance limitado en idiomas no occidentales y por basarse en conocimientos mayormente anglo-europeos. Ciertos tests de veracidad o comprensión (como por ejemplo *TruthfulQA* o *SuperGLUE*) se anclan en información y referencias culturales de Estados Unidos, asumiendo implícitamente que el conocimiento relevante para la "verdad" o la "coherencia" proviene de esa esfera. Esto significa que un modelo de IA podría rendir muy bien según métricas actuales y aun así fallar estrepitosamente en comprender matices culturales o necesidades locales de, digamos, comunidades andinas, rurales africanas o minorías lingüísticas. En pocas palabras, las actuales definiciones de qué constituye una IA "segura" o "avanzada" reflejan puntos de vista parcializados. Esto conlleva el riesgo de que, aunque logremos alinear las IAs con ciertos valores, estemos dejando fuera del cuadro las experiencias y valores de miles de millones de personas.

La naturaleza global del problema de alineación también se hace evidente al considerar la escala de los posibles daños. Un *desalineamiento* grave de una IA poderosa no reconocería fronteras: si por ejemplo un sistema de inteligencia artificial militar escapa al control, su impacto podría desencadenar conflictos internacionales. Del mismo modo, algoritmos financieros desalineados podrían causar estragos económicos en mercados interconectados globalmente. De ahí que expertos como Yoshua Bengio subrayen que la coordinación internacional es tan crucial como la solución técnica en sí: incluso suponiendo que descubramos cómo controlar a una hipotética superinteligencia, de poco serviría si no existen instituciones políticas globales fuertes que impidan el uso abusivo de ese poder por parte de alguna nación o corporación. Bengio sostiene que debemos evitar que "un solo humano, una sola corporación o un solo gobierno" abuse del poder de la IA en detrimento del bien común. En última instancia, alinear la IA con los intereses de la humanidad implica también alinear a los propios humanos entre sí: requiere acuerdos internacionales, tratados y marcos de gobernanza que frenen tanto la carrera armamentista en IA como la tentación de emplearla para dominación política o económica.

Se han dado ya algunos pasos iniciales hacia ese diálogo global. En noviembre de 2023, la Cumbre de Seguridad de la IA de Bletchley Park (Reino Unido) reunió a gobiernos y empresas para discutir riesgos de la IA avanzada, marcando un primer reconocimiento oficial del tema a alto nivel. Le siguieron en 2024 encuentros en Seúl y en febrero de 2025 una cumbre en París enfocada en "IA inclusiva y sostenible". Estas iniciativas vislumbraban la posibilidad de un marco cooperativo internacional. Sin embargo, han enfrentado críticas sustantivas sobre su inclusividad. Observadores señalaron la escasa representación de países en desarrollo y la falta de transparencia en la selección de participantes. Por ejemplo, la cumbre de París (AI Action Summit 2025, impulsada por Francia) decepcionó a muchos representantes del Sur Global. Pese a la retórica de buscar una "IA para las personas y el planeta", en la práctica el evento se percibió más como una

plataforma para posicionar a Francia en la competencia geopolítica de la IA junto a Estados Unidos y China, con la promesa de una inversión de 109 mil millones de euros para hacer del país una potencia en IA. Activistas y académicos del Sur Global allí presentes replicaron la queja de haber sido *instrumentalizados*: se habló de inclusión, pero los recursos y las decisiones permanecieron en manos de los de siempre. En general, la limitada diversidad socioeconómica de las naciones participantes en estas cumbres ha hecho dudar de cuán genuinamente *globales* serán las soluciones propuestas. Si los foros donde se discute la alineación de la IA excluyen a voces fuera del eje EUA-Europa-China, existe un riesgo real de que cualquier acuerdo resultante sea parcial e incompleto.

Un área concreta donde la ausencia de voces diversas es palpable es el establecimiento de normas técnicas y éticas. Como vimos, las métricas para evaluar sistemas de IA han sido anglocéntricas. Afortunadamente, comienza a haber esfuerzos para corregir esta tendencia: investigadores de regiones antes subrepresentadas están desarrollando *benchmarks* adaptados a sus contextos. Se han creado, por ejemplo, conjuntos de datos y evaluaciones para que los modelos manejen mejores tareas en lenguas africanas y del sur de Asia, y se han incorporado idiomas indígenas americanos como el quechua o el criollo haitiano en pruebas de rendimiento de modelos lingüísticos. Estos avances demuestran que es posible enriquecer la noción de "alineación" incorporando dimensiones culturales y lingüísticas variadas. Una IA alineada no solo debe "portarse bien" según estándares universales, sino también *entender* el mundo diverso en el que opera. Y para lograr esto último, es indispensable la participación de comunidades científicas y académicas de todos los continentes.

Por último, cuando hablamos de alineación global conviene recordar que la IA no opera en el vacío: se inserta en estructuras económicas mundiales ya marcadas por la desigualdad. Actualmente, los recursos para desarrollar IA de punta (datos masivos, supercomputación, talento altamente especializado) están concentrados en unas pocas naciones y empresas. Esto genera lo que algunos analistas describen como patrones de dependencia y colonización de datos. Países con menos infraestructura tecnológica corren el riesgo de convertirse en *simples consumidores* de sistemas de IA desarrollados en el exterior, sin capacidad de incidencia en su diseño. Esta dependencia tecnológica no solo tiene implicaciones económicas (por ejemplo, pérdida de competitividad local y fuga de valor agregado), sino también éticas y políticas: ¿qué significa alinear la IA con *nuestros* valores si la mayor parte de las herramientas y plataformas provienen de fuera y traen incorporada la visión de otros? Alinear la IA a nivel global implica entonces enfrentar estas inequidades subyacentes. Requiere construir puentes para que todos los países, especialmente los de menor ingreso, puedan participar en la creación de la IA y no solo en su consumo. Si no, la alineación podría volverse un nuevo campo de batalla de poder, donde aquellos que controlan la tecnología definen las reglas de juego y los valores a considerar, perpetuando una suerte de colonialismo digital.

En síntesis, el desafío de la alineación de la IA es tan global en sus causas y consecuencias que ningún país o región puede quedar al margen de la búsqueda de soluciones. Reconocer esta realidad nos lleva al siguiente paso: preguntarnos cómo el Sur Global, y en particular América Latina, percibe y afronta este desafío, y qué puede aportar al debate mundial.

Desigualdades tecnológicas y la necesidad de la perspectiva del Sur Global

El desequilibrio entre el Norte y el Sur en la era de la IA se manifiesta en múltiples dimensiones. Una de las más evidentes es la brecha económica y de capacidad tecnológica. Como ya se señaló, existe una proyección de que la mayor parte de los beneficios económicos de la IA podría concentrarse en economías desarrolladas. Esto amplificaría un patrón histórico: las revoluciones tecnológicas previas (industrial, informática) inicialmente acentuaron las diferencias de productividad y riqueza entre países centrales y periféricos. Sin intervención deliberada, la revolución de la IA correría la misma suerte o peor, dado que la data y los algoritmos tienden a tener efectos de red y de escala aún más pronunciados. De hecho, salvo contadas excepciones, los países del Sur Global invierten mucho menos en I+D en IA, cuentan con menor infraestructura de supercómputo y producen proporcionalmente menos publicaciones científicas en aprendizaje automático que sus contrapartes del Norte. Para ilustrar, África entera aporta menos del 1% de las publicaciones en las principales conferencias de IA. América Latina, por su parte, si bien cuenta con talentosos grupos de investigación, enfrenta limitaciones presupuestales y fuga de cerebros hacia polos tecnológicos del extranjero. Esta disparidad en la producción de conocimiento e innovación en IA implica que la mayoría de las soluciones de IA desplegadas en países en desarrollo han sido diseñadas y entrenadas con datos de contextos muy distintos.

Ante este panorama, surge una preocupación central desde el Sur Global: cómo evitar una dependencia absoluta de las IA foráneas y las "cajas negras" propietarias. El riesgo de convertirse en “meros consumidores de avances externos” ha sido destacado, por ejemplo, en la estrategia nacional de IA de Argentina. Dicho documento propone esfuerzos coordinados para posicionar al país en un rol más activo y evitar una adopción tardía que comprometa el desarrollo local. No se trata solo de orgullo nacional; está en juego la soberanía tecnológica. Si los algoritmos que determinan decisiones críticas (desde la aprobación de un crédito bancario hasta la distribución de ayudas sociales, o la detección de desinformación en redes) son opacos y vienen empaquetados desde Silicon Valley o Shenzhen, las sociedades receptoras pierden agencia. Además, existe la posibilidad de sesgos sistémicos: algoritmos entrenados principalmente con datos de usuarios y realidades del Norte pueden fallar o cometer injusticias al aplicarse en contextos del Sur. Ya hemos visto casos documentados de sistemas de visión artificial que tienen menor precisión al reconocer rostros de personas de piel más oscura, o asistentes de voz que entienden peores acentos no estándar del inglés. En América Latina, una preocupación concreta es cómo estas herramientas tratan el idioma español (y las variantes locales del mismo), así como las lenguas indígenas. Un sistema de IA alineado con los usuarios latinoamericanos debería, por ejemplo, *entender* correctamente las distintas modalidades del español rioplatense, caribeño, andino, etc., y no replicar sesgos discriminatorios presentes en los datos (por ejemplo, asociar ciertos dialectos o nombres a perfiles de menor ingreso o educación).

La colonización de datos mencionada en informes internacionales alude a una dinámica en la que las naciones desarrolladas y sus corporaciones tecnológicas extraen valor de los datos generados globalmente (incluidos los del Sur) para entrenar sus modelos, pero luego ese valor vuelve en forma de productos y servicios por los que los países periféricos pagan, sin necesariamente compartir los beneficios. Un ejemplo podría ser el uso de abundantes datos de usuarios latinoamericanos en plataformas de internet cuya IA de recomendación se nutre de esas interacciones; la empresa propietaria monetiza ese conocimiento (vía publicidad dirigida, por ejemplo), mientras que los países donde se

originan los datos solo experimentan en muchos casos los efectos negativos (manipulación informativa, polarización, invasión de privacidad) sin retribución equivalente. Esta lógica extractiva recuerda patrones coloniales históricos, de ahí el término y es precisamente lo que muchas voces del Sur Global quieren poner sobre la mesa en cualquier discusión de alineación: ¿Alineada con quién y para beneficio de quién? Una IA verdaderamente alineada con el interés global tendría que minimizar esas dinámicas de dominación y reparto desigual de beneficios.

En foros internacionales, representantes de países en desarrollo han enfatizado que cualquier acuerdo o tratado sobre IA debe atender sus preocupaciones concretas. Por ejemplo, se ha debatido la idea de un tratado global que limite ciertos desarrollos peligrosos de IA (análogamente a los tratados de no proliferación nuclear). Sin embargo, algunos países del Sur han mostrado escepticismo, temiendo que tales acuerdos simplemente congelen el statu quo de desventaja tecnológica. El científico Yoshua Bengio plantea el punto con claridad: *¿por qué firmarían los países del Sur Global un tratado internacional sobre IA?* Su respuesta es que solo lo harían si ese acuerdo garantiza que la IA no sea usada como herramienta de dominación (incluida la dominación económica) y que sus beneficios se compartan globalmente. En otras palabras, para que haya voluntad de cooperar en la contención de riesgos de IA, las naciones en desarrollo necesitan ver que también se abordarán las injusticias y acceso del orden digital actual. Esto podría traducirse en compromisos de transferencia de tecnología, acceso abierto a avances científicos, financiamiento para que los países rezagados desarrollen capacidades propias en IA, y reglas que limiten el uso de la IA para ventajas geopolíticas desleales o explotación económica.

La perspectiva del Sur Global enfatiza, por tanto, la inseparabilidad entre la alineación técnica de la IA y la equidad global. No sería aceptable —ni factible políticamente— pretender que todos los países se sumen a un esfuerzo de control de IA superinteligente, si ese esfuerzo es percibido como impuesto desde arriba o como un pacto entre potencias que deja al resto en posición subordinada. Esto se vincula con un concepto más amplio: la **gobernanza democrática e inclusiva de la IA**. Los países latinoamericanos, africanos y asiáticos han abogado en foros multilaterales porque las discusiones sobre normas y estándares de IA ocurran en espacios donde ellos tengan voz, como Naciones Unidas, evitando arreglos exclusivos de clubes tecnológicos. De hecho, en 2021 todos los Estados miembro de la UNESCO (que incluyen a 193 países de todas las regiones) adoptaron por consenso la primera *Recomendación sobre la ética de la IA*. Este documento, un marco normativo no vinculante pero orientador, es un ejemplo de construcción multilateral inclusiva. Contiene principios sobre transparencia, justicia, privacidad, responsabilidad y promoción del bienestar, acordados con participación activa de países del Sur. La adopción universal de esa Recomendación muestra que **sí es** posible lograr consensos globales cuando todas las partes se sientan en la mesa. Ahora bien, el reto es llevar esos principios a la práctica.

En resumen, las desigualdades tecnológicas existentes hacen imperativo que la perspectiva del Sur Global se integre en el debate de alineación de la IA. No solo porque las naciones en desarrollo demandan legítimamente un reparto más justo de los beneficios y una protección frente a los perjuicios, sino también porque ignorar esas preocupaciones haría inviable cualquier estrategia global. Alinear la IA con la humanidad no puede significar alinear la IA solo con un subconjunto privilegiado de la humanidad. La

siguiente sección profundiza en cómo, específicamente desde América Latina, se están planteando estos temas, con qué matices y propuestas.

Voces latinoamericanas: preocupaciones y prioridades

América Latina, como parte del Sur Global, aporta al debate sobre la IA una voz marcada por las experiencias de la región en materia de desarrollo, desigualdad y derechos humanos. En varios países latinoamericanos se han realizado estudios de opinión y consultas públicas que revelan una ciudadanía alerta tanto a las promesas como a los riesgos de la inteligencia artificial. Un estudio llevado a cabo en 2024 por la fundación Luminate e Ipsos en cuatro países (Argentina, Brasil, Colombia y México) ofrece un interesante panorama: 55% de las personas encuestadas apoyan que la IA esté sujeta a regulación gubernamental. Este respaldo mayoritario a la regulación indica que los latinoamericanos no ven a la IA con ingenuidad techno-optimista, sino que reconocen la necesidad de establecer límites y reglas para su uso. De hecho, el apoyo a la regulación asciende a 65% entre quienes están más familiarizados con las herramientas de IA, lo que sugiere que el conocimiento no genera complacencia sino mayor conciencia de los posibles problemas.

Una de las inquietudes principales es que la IA pueda exacerbar las desigualdades sociales preexistentes. En la región más desigual del planeta, un 37% de los encuestados expresó el temor de que las aplicaciones de IA terminen aumentando la brecha entre grupos sociales. Este sentir no es infundado: si las soluciones de IA no están diseñadas pensando en contextos de inequidad, podrían, por ejemplo, concentrar aún más riqueza en manos de industrias altamente tecnificadas dejando atrás a trabajadores menos calificados, o podrían facilitar la focalización de servicios solo hacia segmentos rentables de la población, ignorando a comunidades marginadas. Los latinoamericanos también han vivido de cerca los efectos de la tecnología digital desregulada en la última década, especialmente a través de las redes sociales. No es casual que un portavoz regional como Felipe Estefan (director en América Latina de Luminate) haya afirmado que *con la IA tenemos la oportunidad de aprender de los errores cometidos con las plataformas de redes sociales*. La falta de rendición de cuentas de estas últimas derivó en problemas palpables en la región: difusión incontrolada de desinformación, propagación de discursos de odio, aumento de la polarización política y una preocupante vigilancia masiva sobre la ciudadanía. La lección es clara: no se debe esperar a que la IA cause estragos similares para actuar. Hay un consenso emergente en América Latina de que la ética y la regulación proactiva deben acompañar el despliegue de la IA desde el principio, para evitar repetir las dinámicas nocivas ya vistas en el entorno digital.

Otro tema sensible para la opinión pública latinoamericana es el uso de IA en decisiones que afectan derechos o bienes públicos fundamentales. La mencionada encuesta reveló fuertes niveles de oposición a delegar ciertas decisiones en manos de algoritmos: 54% de los participantes rechazó que se utilice IA para tomar decisiones en tribunales de justicia, 51% se opuso a emplearla para redactar leyes, y 50% vio inaceptable que determine quién tiene derecho a recibir beneficios sociales del Estado. Estas cifras indican una saludable precaución frente a la automatización de funciones públicas. Detrás de ellas subyacen preocupaciones éticas: la justicia, por ejemplo, debe ser *personalizada y con garantías humanas*, por lo que un juez artificial despierta escepticismo sobre su capacidad de ponderar las complejidades de cada caso y mantener imparcialidad sin sesgos algorítmicos. De igual modo, dejar en manos de una IA decisiones legislativas o de

asignación de ayudas sociales despierta recelo a que se pierda la empatía, la deliberación democrática y la consideración de circunstancias individuales. En resumen, para muchos latinoamericanos la IA no debe reemplazar el juicio humano allí donde esté en juego la dignidad o los derechos de las personas. Más bien, se la ve como una herramienta que debe complementar al humano, bajo su control y con transparencia.

Estas actitudes ciudadanas han ido permeando la agenda política en varios países de la región. Durante 2024, se vieron avances notables en iniciativas de regulación de la IA. Un caso destacado es Brasil, cuyo Senado aprobó un proyecto de ley para una gobernanza responsable de la IA. Dicho proyecto se inspiró en parte en la propuesta de Ley de IA de la Unión Europea, adaptando un enfoque de clasificación de riesgos: prohibir aplicaciones de IA con riesgos inaceptables para los derechos fundamentales y exigir evaluaciones de impacto para sistemas de alto riesgo. El texto brasileño deja claro que no se trata de elegir entre innovación y derechos, sino de equilibrarlos, afirmando que no existe un dilema excluyente entre la protección de derechos humanos y el desarrollo económico basado en IA. Este movimiento legislativo en Brasil, la mayor economía latinoamericana, es simbólico del creciente activismo regulatorio en la región. Países como Chile, Colombia y México también han iniciado discusiones parlamentarias o elaborado directrices éticas sobre IA. Por su parte, Uruguay y Argentina contaron tempranamente con estrategias nacionales de IA (en 2019, fueron de los primeros fuera del mundo desarrollado en formular planes nacionales), centradas en la adopción responsable y la capacitación de talento local.

Sin embargo, la región enfrenta desafíos importantes para llevar del dicho al hecho estas buenas intenciones. Un obstáculo común es el poder de lobby de las grandes tecnológicas en contra de regulaciones estrictas. En el proceso brasilero, investigaciones revelaron que un 31% de los espacios de debate en audiencias públicas fue ocupado por representantes de empresas, superando la participación de academia (26%) y sociedad civil (19%). Además, se detectó que a veces los voceros corporativos se presentaban como si representaran otros sectores, intentando diluir la percepción de sus intereses particulares. Esto demuestra el desafío de regular la IA en entornos donde las asimetrías de poder son marcadas: los Estados latinoamericanos, con recursos limitados, deben negociar con corporaciones multinacionales cuyo alcance y conocimiento técnico a menudo excede el de las agencias públicas locales. Para contrarrestar esto, la colaboración regional y el aprendizaje de las experiencias de otros países resultan cruciales. Por ejemplo, Brasil ha tomado inspiración de Europa; del mismo modo, otros países de la región observan de cerca el caso brasileño para replicar buenas prácticas y evitar tropiezos.

A pesar de las disparidades internas, América Latina exhibe también historias de éxito y capacidades destacables en el ámbito de la IA. Un informe de la CEPAL que construyó el Índice Latinoamericano de Inteligencia Artificial (ILIA) situó en 2024 a Chile, Brasil y Uruguay como los países con mayor desarrollo relativo en IA en la regiónelpais.com. Según este índice, estas naciones no solo han avanzado en la implementación de tecnologías IA en diversos sectores, sino que han orientado sus estrategias nacionales para expandir el uso de IA en toda la economía y la sociedad, disponiendo de entornos favorables para la investigación e innovación en IAelpais.com. El caso de Uruguay es particularmente llamativo: a pesar de su tamaño pequeño, ha logrado posicionarse como un *hub* digital regional, con alta penetración de internet, programas educativos pioneros

(como el Plan Ceibal) y un gobierno electrónico robusto. Uruguay publicó una primera Estrategia de IA en 2019 y actualmente (2024-2025) actualiza dicha estrategia con miras a 2030. En la visión uruguaya, plasmada en documentos oficiales, se enfatiza que la IA debe desarrollarse con un enfoque en derechos humanos, equidad social y sostenibilidad, y que el país aspira a integrar la cooperación internacional en IA desde una perspectiva latinoamericana. Uruguay ha destacado en foros globales la necesidad de que la gobernanza internacional de la IA tome en cuenta las particularidades de América Latina y el Caribe, para asegurar que la región tenga acceso equitativo a las oportunidades y beneficios que la IA puede brindar. Este llamado uruguayo resuena con las posturas de otros países vecinos, evidenciando una identidad latinoamericana común en esta materia.

El enfoque latinoamericano hacia la IA, en síntesis, tiende a ser humanista y orientado al desarrollo inclusivo. A diferencia de algunas narrativas dominantes en Estados Unidos, por ejemplo, donde la discusión sobre alineación a veces se centra en futuribles apocalípticos (IA superinteligente rebelde), en América Latina la conversación suele aterrizar en preocupaciones más inmediatas: cómo garantizar que la IA no reproduzca las injusticias sociales, cómo aprovecharla para el bien público (salud, educación, lucha contra la pobreza) y cómo regularla democráticamente. Esto no significa que la región ignore los riesgos a largo plazo como los existenciales; de hecho, países como México y Brasil han participado activamente en espacios de la ONU discutiendo esos riesgos. Pero sí implica que, en la priorización local, la IA ética se entrelaza con agendas de derechos digitales, reducción de brechas y mejora de la calidad de vida. En un sentido amplio, podríamos decir que para América Latina *alinear la IA* equivale a alinearla con los objetivos de desarrollo sostenible y con la protección de los derechos humanos.

Un claro ejemplo de la perspectiva proactiva de la región es la integración de la IA con la agenda de desarrollo. La mencionada estrategia argentina de IA subraya que su fin último es alcanzar resultados significativos en los objetivos de desarrollo nacional, vinculados a los Objetivos de Desarrollo Sostenible (ODS) de Naciones Unidas. Esto refleja una visión donde la IA es un medio para un fin mayor (el desarrollo humano sostenible), no un fin en sí misma. Asimismo, la UNESCO, cuya oficina regional en Santiago de Chile ha trabajado estrechamente con los gobiernos latinoamericanos, enfatiza que con la IA se abre una *"oportunidad única para la transformación productiva, la innovación educativa y el fortalecimiento institucional"* en la región, siempre y cuando se construya una cooperación significativa entre los países y se establezcan marcos de gobernanza efectivos y éticos que garanticen que los beneficios lleguen a toda la población. Nótese la importancia otorgada a la cooperación regional: América Latina parece consciente de que solo actuando en bloque y compartiendo conocimientos podrá tener peso en la arena global de la IA. De hecho, ya existen redes y foros intrarregionales, como el Foro Latinoamericano de Ética de la IA, la iniciativa IA Latinoamérica impulsada por BID y CEPAL, o encuentros ministeriales coordinados por UNESCO y OEA, donde los países intercambian experiencias en regulación, educación en IA y desarrollo de capacidades.

En conclusión, las voces latinoamericanas en el debate sobre la alineación de la IA ponen sobre la mesa la centralidad de la justicia social, la responsabilidad democrática y la búsqueda del desarrollo compartido. Estas preocupaciones y prioridades, arraigadas en la realidad de nuestras sociedades, complementan y enriquecen la discusión global sobre la IA de manera indispensable. Incorporarlas no solo responde a un imperativo

moral de inclusión, sino que mejora las probabilidades de encontrar soluciones robustas y legítimas para que la IA trabaje en favor de *toda* la humanidad.

Iniciativas y estrategias latinoamericanas para una IA alineada

El interés de América Latina por influir en la trayectoria de la IA no se queda en el plano discursivo; se refleja también en acciones concretas de política pública, en el desarrollo de capacidades locales y en aportes a marcos normativos internacionales. Como se mencionó, varios países de la región se adelantaron tempranamente con estrategias nacionales de inteligencia artificial. México fue pionero publicando una estrategia en 2018, seguida en 2019 por Argentina, Uruguay, Chile y Colombia, y en 2020-2021 por Brasil, Perú y otros. Esta ola de estrategias nacionales indica que los gobiernos latinoamericanos reconocieron la importancia estratégica de la IA y buscaron articular una visión propia. En la práctica, dichas estrategias suelen abordar tres ejes fundamentales: (1) Gobernanza y marco regulatorio (principios éticos, leyes, agencias de supervisión), (2) Desarrollo de capacidades (formación de talento, inversión en I+D, infraestructuras como centros de supercómputo) y (3) Impulso a la adopción en sectores productivos y gubernamentales. Aunque la implementación ha sido desigual y enfrenta limitaciones de recursos, la existencia de estos planes es ya un paso adelante en alinear la IA con prioridades nacionales.

Tomemos el caso de Uruguay, cuya Estrategia Nacional de IA (actualizada al período 2024–2030) ofrece elementos ilustrativos. En su introducción, el documento reconoce que las repercusiones económicas, sociales, culturales y ambientales de la IA "no se distribuyen equitativamente entre los países, ni dentro de ellos", alertando así sobre la brecha global y doméstica. Uruguay enmarca su estrategia dentro de esfuerzos internacionales a los que *"ha contribuido activamente"* para establecer bases de una *"gobernanza de la IA global e inclusiva"*, buscando marcos comunes e interoperables que garanticen un desarrollo y uso ético, seguro y responsable de la IA en beneficio de la humanidad. Esta explícita alineación con la agenda global ética (Uruguay hace referencia directa a la Recomendación de la UNESCO sobre ética de la IA) va de la mano de un énfasis regional: *"junto con los países de nuestra región, Uruguay ha destacado la necesidad de que la gobernanza internacional [...] contemple las particularidades de América Latina y el Caribe"*. Se busca así fortalecer el acceso equitativo de la región a las oportunidades de la IA. A nivel interno, la estrategia uruguaya propone políticas públicas co-construidas con múltiples actores (gobierno, sector privado, academia, sociedad civil), y prioriza el desarrollo sostenible. En suma, Uruguay persigue una IA alineada con derechos humanos, integrada en la cooperación internacional pero atenta a las realidades locales, y orientada al bien social. Es un modelo interesante de cómo un país pequeño puede posicionarse éticamente y contribuir a la conversación global.

Por su parte, Argentina elaboró el *Plan Nacional de IA* con un enfoque integral similar. En su prólogo, el plan argentino resalta la dualidad de oportunidades y riesgos: la IA ofrece "oportunidades únicas" para el crecimiento y para resolver problemas complejos de la humanidad, pero a la vez obliga a enfrentar numerosos desafíos que pueden poner en riesgo derechos y vulnerar libertades, además de ampliar brechas entre países y dentro de ellos. La respuesta estratégica de Argentina fue un abordaje multi-actoral (involucrando gobierno, sector productivo, sistema científico-tecnológico, academia, sociedad civil e incluso organismos internacionales) y orientado a maximizar el beneficio social de la IA. Un aspecto notable es la insistencia en que Argentina debe "posicionarse

rápidamente en un rol preeminente" frente a esta tecnología, evitando quedar rezagada. Para ello, plantea formar talento local, fomentar la innovación autóctona y estimular la adopción de IA en sectores estratégicos de la economía. El plan vincula explícitamente sus metas con los Objetivos de Desarrollo Sostenible (ODS) de la ONU, subrayando que la IA debe servir para *aumentar las capacidades humanas* y promover el desarrollo del país. Esto refleja, nuevamente, la visión teleológica de alinear la IA con fines de desarrollo humano. Cabe señalar que la implementación de este plan sufrió altibajos por cambios de gobierno y coyunturas económicas difíciles, pero estableció un marco conceptual valioso que seguramente seguirá informando políticas futuras.

Otros países han adoptado enfoques interesantes también. Chile creó en 2021 una Política Nacional de IA con fuerte énfasis en la dimensión ética y en la inserción de Chile en redes globales de investigación. Colombia, a través de su estrategia de 2019, hizo hincapié en el uso de IA para mejorar servicios públicos y promover emprendimientos locales en IA. Brasil, además de la legislación en trámite, cuenta con una Estrategia de IA (2021) que dedica capítulos a la capacitación masiva de profesionales en tecnologías de IA y a la atracción de inversiones privadas al ecosistema local. En Centroamérica, países más pequeños como Costa Rica y Uruguay han colaborado en proyectos piloto de IA aplicada a la agricultura sostenible y a la gestión ambiental, mostrando que la IA puede alinearse con agendas ecológicas y comunitarias.

A nivel supranacional, América Latina también está tratando de articular una voz unida. La Comisión Económica para América Latina y el Caribe (CEPAL) ha asumido un rol activo, brindando asistencia técnica para estrategias nacionales (incluyendo una metodología de índices de preparación como ILIA) y propiciando diálogos regionales. Un evento importante fue la conferencia de alto nivel "IA en América Latina y el Caribe: retos, estrategias y gobernanza para el desarrollo regional", organizada en marzo de 2025 en Santiago de Chile. Allí, líderes gubernamentales, expertos, sector privado y sociedad civil de toda la región se reunieron para discutir cómo usar la IA como motor de desarrollo sostenible con equidad y transparencia. La UNESCO presentó en esa ocasión su conjunto de herramientas, incluyendo la ya citada Recomendación Ética (2021) y la Metodología de Evaluación de Preparación (RAM), esta última diseñada para diagnosticar qué tan listos están los países para una IA ética y responsable. Gracias a la aplicación del RAM, 14 países (varios latinoamericanos) han podido diseñar o actualizar sus estrategias nacionales con apoyo de la UNESCO. Esto evidencia que la cooperación internacional, bien orientada, está aportando insumos valiosos a la región. Adicionalmente, la conferencia abordó la creación de una agenda regional de gobernanza de IA y el diálogo con el sector privado para promover buenas prácticas en el uso de la tecnología.

Un punto recurrente en las iniciativas latinoamericanas es la educación y sensibilización sobre IA. Conscientes de que se requiere no solo marcos legales sino también una ciudadanía empoderada digitalmente, varios gobiernos han lanzado cursos en línea, campañas de difusión e incluso ajustes curriculares para incluir nociones de IA en la educación formal. Uruguay, por ejemplo, integró contenidos de pensamiento computacional e IA en programas de educación secundaria. Argentina y Brasil han apoyado hackatones y desafíos para jóvenes orientados a soluciones de IA con impacto social. Estas actividades buscan alinear el *uso* de la IA con valores de innovación abierta, inclusión y bienestar público, formando a la próxima generación de desarrolladores y usuarios con ese ethos.

Finalmente, es importante mencionar la contribución de academia y sociedad civil latinoamericanas al debate global. En la academia, centros como el Instituto de IA de la Universidad de Buenos Aires, el Laboratorio de IA del Tecnológico de Monterrey, o grupos en la Universidade de São Paulo, están publicando investigaciones tanto técnicas como sobre impactos socio-éticos de la IA. Investigadores latinoamericanos participan en comités de ética de organizaciones internacionales (por ejemplo, la COMEST de la UNESCO cuenta con expertos de la región). Desde la sociedad civil, organizaciones como Derechos Digitales (con presencia en varios países), Fundación Karisma (Colombia), Observatorio de IA (Argentina), entre otras, monitorean los desarrollos de IA, denuncian potenciales abusos (como sistemas de vigilancia biométrica sin controles) y proponen guías de buenas prácticas. Estas entidades han llevado la voz latina a espacios globales como el Foro de Gobernanza de Internet (IGF) y eventos de AI for Good de la ONU, abogando por enfoques de IA centrados en el ser humano y en los derechos.

En suma, las estrategias e iniciativas de América Latina conforman un mosaico de esfuerzos por alinear la IA con las necesidades locales y al mismo tiempo incidir en las reglas globales. Hay conciencia de los límites (ningún país latinoamericano por sí solo cambiará el rumbo de la IA mundial), pero también hay convicción de que actuando conjuntamente y aportando sus visiones, la región puede ayudar a reorientar la tecnología hacia fines más humanos y justos. Este activismo pragmático —desde leyes nacionales hasta participación en estándares internacionales— demuestra que la perspectiva del Sur Global no solo critica, sino que construye alternativas y soluciones concretas. Incorporar dichas soluciones al mainstream global podría beneficiar a todos, al diversificar las aproximaciones al problema de alineación y enriquecer las herramientas con que contamos para enfrentarlo.

Hacia una gobernanza global inclusiva de la IA

Considerando todo lo anterior, resulta claro que la alineación de la IA debe concebirse como un proyecto global inclusivo, donde las necesidades y visiones del Sur Global tengan un lugar central. ¿Cómo avanzar hacia tal modelo de gobernanza global? A continuación, se proponen, a modo de *propuesta fundamentada*, algunos lineamientos clave extraídos del análisis precedente:

- **Equidad en los beneficios y prevención de la dominación:** Cualquier acuerdo o marco internacional sobre IA (ya sea un tratado vinculante, declaraciones de principios o estándares técnicos) debe incorporar cláusulas que aseguren la distribución global de los beneficios de la IA y eviten su uso como instrumento de dominación. Esto implica, por ejemplo, compartir conocimientos y transferir tecnologías a países en desarrollo, apoyar la creación de capacidad local (formación de expertos, infraestructuras computacionales) y prohibir el uso de sistemas de IA por parte de estados o empresas para prácticas neocoloniales (vigilancia masiva de poblaciones vulnerables, manipulación informativa en otros países, etc.). Sin estas garantías, difícilmente las naciones del Sur verán atractivo sumarse a iniciativas globales de control de IA, ya que las percibirían sesgadas a favor de los poderosos.
- **Participación plural en la formulación de normas:** Los foros que delinean las normas de la IA (sean sobre seguridad, ética o métricas de desempeño) deben ser intrínsecamente plurales. Esto significa reforzar espacios multilaterales como

Naciones Unidas y la UNESCO, asegurando representación de países de todos los continentes en los grupos de trabajo y comités especializados. Asimismo, adoptar enfoques *multistakeholder* (múltiples actores): incluir, además de gobiernos, a la sociedad civil global, academia de distintas regiones, comunidades indígenas y sector privado local. Por ejemplo, en la posible creación de un *Consejo Mundial de IA*, se debería reservar asientos no solo para las grandes potencias sino también para consorcios regionales del Sur. La diversidad de voces garantizará que las definiciones de "IA segura" o "IA beneficiosa" resultantes tengan legitimidad y reflejen una perspectiva verdaderamente global.

- **Marcos éticos universales con flexibilidad cultural:** Documentos como la Recomendación Ética de la UNESCO (2021) proveen un buen punto de partida al establecer principios universales acordados, pero es crucial que dichos principios se aterricen de manera contextualizada en cada región. Un enfoque recomendable es adoptar principios comunes, pero con implementación localmente adaptable. Por ejemplo, el principio de transparencia es universal, pero cómo se garantiza puede variar según las capacidades e instituciones de cada país. Del mismo modo, el respeto a la diversidad cultural puede requerir que ciertos usos de IA (como generación de contenido) sean supervisados por comunidades locales para evitar representaciones ofensivas o inapropiadas. La alineación global no debe borrar las particularidades; más bien debe proporcionar un marco donde estas convivan. En última instancia, se busca una IA alineada con los valores humanos en plural, entendiendo por ello un núcleo de derechos humanos compartidos (dignidad, justicia, libertad) y la riqueza de expresiones culturales diversas que los contextualizan.

- **Sistemas de evaluación inclusivos y responsables:** Es imperativo desarrollar *benchmarks* y evaluaciones de IA que incluyan la diversidad global, como ya se ha empezado a hacer incluyendo idiomas subrepresentados. Una iniciativa concreta podría ser crear un Consorcio Global de Evaluación de IA con sede rotativa en diferentes regiones, que diseñe pruebas multilingües y multiculturales para las IAs más avanzadas. Este consorcio debería involucrar universidades y centros de investigación del Sur Global, para aprovechar su conocimiento del contexto local. De igual forma, se necesita incorporar enfoques de auditoría algorítmica independiente en todos los países: equipos locales capaces de auditar sistemas de IA importados, para verificar su alineación con criterios éticos locales. Esto requerirá financiamiento y cooperación para formar auditores en países que hoy no cuentan con ellos.

- **Fortalecimiento de capacidades y reducción de brechas:** Ninguna gobernanza global será efectiva si persisten enormes disparidades de capacidad entre países. Por ende, una pieza fundamental de la agenda debe ser un plan de desarrollo de capacidades en IA para países de renta baja y media. Esto podría gestionarse a través de organismos internacionales (Banco Mundial, ONU) creando fondos dedicados a: formación de expertos en IA en universidades del Sur (becas, intercambios), instalación de centros de supercómputo regionales accesibles, traducción de contenidos formativos a múltiples idiomas, y soporte a empresas emergentes locales de base tecnológica. La idea es habilitar a cada país a ser co-creador de soluciones de IA. Al reducir la dependencia tecnológica, también se

facilita que cada nación alinee la IA con sus prioridades nacionales sin imponer modelos ajenos. Como dice el adagio, *dar voz también implica dar poder*; si queremos que el Sur Global tenga voz en la alineación, debe tener poder técnico para desarrollar y modificar sistemas según sus necesidades.

- **Mecanismos de rendición de cuentas globales:** Para asegurar que los compromisos se cumplan, habría que instaurar mecanismos de monitoreo y seguimiento internacionales. Por ejemplo, un panel de reporte anual sobre IA y equidad global, donde se evalúe qué tanto se ha avanzado en inclusión de países rezagados, en reducción de sesgos culturales en IA, en cooperación internacional real (no solo declarativa). Este panel podría funcionar al estilo del IPCC (el panel climático), congregando evidencia de todas las regiones, identificando brechas y recomendando acciones. Asimismo, se podrían establecer indicadores globales de alineación: no solo medir cuántos sistemas han evitado fallas catastróficas, sino también cuán distribuidos están los beneficios de la IA (por ejemplo, porcentajes de PIB regional atribuibles a IA, para ver si sube en el Sur), o cuántos países aplican principios éticos en sus leyes de IA. Teniendo datos y métricas, la comunidad internacional puede ejercer presión y reconocer logros, creando una dinámica de *accountability* (rendición de cuentas) en este ámbito.
- **Enfoque precautorio y cooperativo en investigación avanzada:** Respecto a las fronteras más avanzadas de IA (como proyectos hacia AGI o IA general), es vital que no se transformen en otra carrera armamentista bilateral. Se debería promover la creación de consorcios internacionales para la investigación de IA segura, invitando a participar también a científicos del Sur. Un ejemplo para considerar es la idea de un laboratorio global de IA bajo auspicio público-mundial, donde los avances hacia AGI se conduzcan con transparencia y controles multilaterales. Así se evitaría que solo un país o corporación controle un posible salto disruptivo. Este laboratorio podría ubicarse en un país neutral o rotar sedes, incorporando talentos de diversas procedencias (incluir a jóvenes investigadores latinoamericanos, africanos, asiáticos formados en los programas mencionados). De esta manera, la misma creación científica de la IA general estaría alineada desde su génesis con un espíritu de colaboración global, en vez de competencia de suma cero. Ello redundaría en más seguridad y confianza para todos.

En última instancia, el objetivo de todos estos lineamientos es plasmar en la gobernanza real el principio de que la IA debe convertirse en un catalizador de progreso compartido, y no en un factor que profundice desigualdades. Como mencionó un funcionario de la UNESCO, para lograrlo hay que fortalecer capacidades en todos los niveles (desde gobiernos hasta ciudadanos) de modo que la IA no sea patrimonio de unos pocos sino una herramienta al servicio de todos. La alineación de la IA no solo concierne a evitar riesgos extremos; también significa alinear la tecnología con los anhelos universales de desarrollo humano, justicia y dignidad.

Conclusiones

La alineación de la inteligencia artificial es, a la vez, un desafío técnico, ético y político de alcance global. Hay que asegurar que las IA del presente y del futuro actúen en coherencia con los valores y el bienestar humanos requiere un esfuerzo mancomunado de la comunidad internacional. En este ensayo hemos sostenido que tal esfuerzo solo será efectivo y legítimo si la perspectiva del Sur Global ocupa un lugar central en la discusión y en las soluciones implementadas. La mirada de América Latina y otras regiones en desarrollo aporta sensibilidades y prioridades imprescindibles: la urgencia de no agravar desigualdades históricas, la demanda de un reparto equitativo de beneficios, la insistencia en marcos éticos construidos desde la diversidad cultural, y la visión de la IA como herramienta para el desarrollo sostenible y no como fin en sí mismo.

Hemos visto cómo, por un lado, los debates y acciones dominantes en torno a la IA han tendido a ser occidental-céntricos, con cumbres internacionales inicialmente poco inclusivas y estándares que no reflejan plenamente la pluralidad del mundo. Esto conlleva serios riesgos de soluciones parcializadas o imposición de una nueva hegemonía digital. Por otro lado, América Latina y el Sur Global no han permanecido pasivos: han articulado preocupaciones legítimas, han propuesto principios (como compartir los frutos de la IA globalmente y prohibir su uso opresivo), y han emprendido iniciativas locales y regionales para alinear la IA con los derechos humanos y el bien común. Países latinoamericanos como Uruguay y Argentina sirven de ejemplo al integrar consideraciones éticas y de equidad en sus estrategias nacionales, a la vez que claman por mayor voz en la gobernanza internacional. La ciudadanía latinoamericana apoya mayoritariamente la regulación de la IA y teme con razón sus posibles impactos en sociedades desiguales, lo que ha impulsado debates democráticos sobre cómo guiar esta tecnología antes de que sea tarde.

En la encrucijada actual, se abren dos futuros posibles. En uno, la IA profundiza las divisiones: los países que hoy lideran continúan acumulando ventaja, imponiendo sus valores por defecto en sistemas usados globalmente, mientras el resto del mundo padece una doble exclusión (no participa en la creación y sufre las consecuencias de la utilización). En otro futuro —el que abogamos por construir— la IA se convierte en un proyecto verdaderamente colectivo, donde todas las naciones, grande o pequeñas, tienen voz en cómo debe usarse esta poderosa herramienta para mejorar la vida en el planeta. Alcanzar este segundo escenario exigirá mucho diálogo, voluntad política, generosidad y quizás nuevas fórmulas institucionales que hoy apenas esbozamos. Pero los primeros pasos ya se están dando: desde la adopción universal de la ética de la IA en UNESCOunesco.org, hasta la colaboración Sur-Sur en investigación y la elevación del tema en la agenda de organismos como el G77+China (donde en 2023 se discutió ciencia, tecnología e IA con énfasis en desarrollo).

En definitiva, alinear la IA con la humanidad significa, inexorablemente, alinear la IA con *toda* la humanidad en su rica diversidad. Cualquier aproximación más limitada corre el riesgo de ser insuficiente o injusta. Como humanidad, enfrentamos un desafío común ante la IA avanzada; pero no partimos del mismo lugar ni con las mismas herramientas. Reconocer esas asimetrías y trabajar activamente para reducirlas es parte integral de resolver el problema de alineación. Una IA que refleje únicamente las prioridades de Silicon Valley o de Beijing difícilmente será una IA alineada con los intereses del granjero andino, de la médica rural africana o del estudiante del Caribe. Por ello, incorporar las

voces de América Latina, de África, de toda la mayoría global, no es un acto de cortesía sino de pragmatismo y sabiduría colectiva.

La apuesta es alta. Si logramos una gobernanza inclusiva, podríamos desbloquear el enorme potencial de la IA para erradicar enfermedades, democratizar la educación, optimizar la agricultura sostenible y tantas otras metas, de manera que esos beneficios alcancen a miles de millones y no solo a unos pocos. Y a la vez, podríamos dormir más tranquilos sabiendo que sistemas cada vez más inteligentes operan bajo salvaguardas y valores compartidos. Si fracasamos, podríamos deslizarnos hacia un mundo de nuevas brechas, conflictos por la supremacía tecnológica y quizás, eventualmente, perder el control sobre criaturas digitales que no entienden ni respetan lo que nos hace humanos.

Este ensayo ha presentado argumentos y evidencias para fundamentar la primera opción: un camino de cooperación global con el Sur Global como protagonista en el debate de la alineación de la IA. Queda como trabajo futuro profundizar en mecanismos específicos para materializar estos principios, así como mantener un monitoreo crítico de cómo evoluciona la situación. En palabras sencillas, se trata de que la próxima fase de la inteligencia artificial esté guiada por la inteligencia colectiva de la humanidad entera. Solo así, la alineación de la IA dejará de ser un problema angustiante para convertirse en una oportunidad de *alianza global* hacia un futuro mejor.

Referencias

Agencia de Gobierno Electrónico y Sociedad de la Información de Uruguay. (2024). *Estrategia Nacional de Inteligencia Artificial del Uruguay 2024–2030*. https://www.gub.uy/agencia-gobierno-electronico-sociedad-informacion-conocimiento/

Bengio, Y. (2024, 12 de enero). *Reasoning through arguments against taking AI safety seriously* [Entrada de blog]. https://yoshuabengio.org/2024/01/12/reasoning-against-ai-safety/

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Brookings Institution. (2023). *AI governance: A critical global challenge*. https://www.brookings.edu/articles/ai-governance-critical-global-challenge/

CAF - Banco de Desarrollo de América Latina y el Caribe. (2023, octubre). *Inteligencia artificial en América Latina: Situación actual y perspectivas*. https://www.caf.com/es/actualidad/noticias/2023/10/inteligencia-artificial-en-america-latina-situacion-actual-y-perspectivas/

CIDOB. (2024). *El Sur Global y la inteligencia artificial: ¿ruptura o integración?* [Infografía]. *Anuario Internacional CIDOB 2025*. https://www.cidob.org/publicaciones/infografias-el-sur-global-y-la-inteligencia-artificial-ruptura-o-integracion

Comisión Económica para América Latina y el Caribe. (2024). *Índice Latinoamericano de Inteligencia Artificial (ILIA)*. https://www.cepal.org/es/publicaciones/48513-indice-latinoamericano-inteligencia-artificial-ilia

Comisión Económica para América Latina y el Caribe. (2025, 4 de marzo). *La inteligencia artificial está transformando al mundo y América Latina y el Caribe no puede quedarse atrás* [Comunicado de prensa]. https://www.cepal.org/es/comunicados/inteligencia-artificial-transforma-mundo-america-latina-caribe-no-puede-quedarse-atras

Comisión Económica para América Latina y el Caribe & Centro Nacional de Inteligencia Artificial. (2024). *Índice Latinoamericano de Inteligencia Artificial (ILIA) 2024* (2.ª ed.). CEPAL/CENIA.

Duque Ruíz, S. (2023). *La revolución silenciosa: Inteligencia artificial, justicia social y cambio comunitario*. *Diálogos – Naturaleza y Sociedad*, (13). Universidad de los Andes. https://openscience.uniandes.edu.co

Estefan, F. (2024, 21 de abril). *Latinoamérica no debe repetir los errores cometidos con las redes sociales frente al avance de la IA*. El País. https://elpais.com/tecnologia/2024-04-21/latinoamérica-no-debe-repetir-los-errores-cometidos-con-las-redes-sociales-frente-al-avance-de-la-ia.html

Larsen, B. C., & Dignum, V. (2024, 22 de octubre). *Cómo alinear la inteligencia artificial con los valores humanos*. Foro Económico Mundial. https://es.weforum.org/agenda/2024/10/como-alinear-la-ia-con-valores-humanos/

Luminate Foundation & Ipsos. (2024). *DemocracIA: Encuesta sobre inteligencia artificial en América Latina*. https://luminategroup.com/

Mohamed, S., Png, M.-T., & Isaac, W. (2020). *Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence*. *Philosophy & Technology, 33*(4), 659–684. https://doi.org/10.1007/s13347-020-00405-8

Naciones Unidas, Asamblea General. (2024a, 21 de marzo). *Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development* [Resolución A/78/L.49]. https://press.un.org/en/2024/ga12588.doc.htm

Naciones Unidas, Asamblea General. (2024b, 25 de junio). *Enhancing international cooperation on capacity-building of Artificial Intelligence* [Resolución]. https://pam.int/un-resolution-artificial-intelligence/

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2021). *Recomendación sobre la ética de la inteligencia artificial*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000380455_spa

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2024, 6 de febrero). *Foro Global sobre la Ética de la IA 2024 – Cambiando el panorama de la gobernanza de la IA* [Comunicado]. UNESCO.

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, CAF & Agesic. (2024). *Segunda Cumbre Ministerial sobre la Ética de la IA en América Latina y el Caribe – Declaración de Montevideo*. UNESCO/CAF/Agesic.

Organización para la Cooperación y el Desarrollo Económicos. (2019). *Principios de la OCDE sobre la inteligencia artificial*. https://oecd.ai/en/dashboards/policy-initiatives

PwC. (2017). *Sizing the prize: What's the real value of AI for your business and how can you capitalise?*. https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html

Rahimi Midani, A. (2025, 7 de julio). *¿Quién posee el futuro?* Delfino.cr. https://delfino.cr/2025/07/quien-posee-el-futuro

República Argentina. (2019). *Plan Nacional de Inteligencia Artificial*. Consejo Económico y Social de la Presidencia de la Nación.

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking.

Tegmark, M. (2017). *Life 3.0: Being human in the age of artificial intelligence*. Alfred A. Knopf.

Zuazo, N. (2025, 21 de marzo). *América Latina ante la IA: ¿regulación o dependencia tecnológica?* El País. https://elpais.com/america-futura/2025-03-21/america-latina-ante-la-ia-regulacion-o-dependencia-tecnologica.html