

HURI-AGE

Red Tiempo de los Derechos



Papeles el tiempo de los derechos

**DERECHOS HUMANOS COMO BIEN PÚBLICO
GLOBAL: INTELIGENCIA ARTIFICIAL Y
MECANISMOS DE CONTROL DEL CUMPLIMIENTO
CON INDEPENDENCIA OPERATIVA.**

**HUMAN RIGHTS AS A GLOBAL PUBLIC GOOD:
ARTIFICIAL INTELLIGENCE AND MECHANISMS FOR
MONITORING COMPLIANCE WITH OPERATIONAL
INDEPENDENCE**

Alberto Delfín Arrufat Cárdua

Prof. Derecho Internacional Público Universitat
Universitat Jaume I de Castellón

Palabras Clave: derechos humanos, derecho del niño, inteligencia artificial.

Key Words: human rights, rights of the child, artificial intelligence

Comité Evaluador de los Working Papers “El Tiempo de los Derechos”

María José Añón (Universidad de Valencia)
María del Carmen Barranco (Universidad Carlos III)
María José Bernuz (Universidad de Zaragoza)
Rafael de Asís (Universidad Carlos III)
Eusebio Fernández (Universidad Carlos III)
Andrés García Inda (Universidad de Zaragoza)
Cristina García Pascual (Universidad de Valencia)
Miguel A. Ramiro (Universidad de Alcalá)
María José González Ordovás (Universidad de Zaragoza)
Jesús Ignacio Martínez García (Universidad of Cantabria)
Antonio E Pérez Luño (Universidad de Sevilla)
Miguel Revenga (Universidad de Cádiz)
Maria Eugenia Rodríguez Palop (Universidad Carlos III)
Eduardo Ruiz Vieytez (Universidad de Deusto)
Jaume Saura (Instituto de Derechos Humanos de Cataluña)

**DERECHOS HUMANOS COMO BIEN PÚBLICO
GLOBAL: INTELIGENCIA ARTIFICIAL Y
MECANISMOS DE CONTROL DEL CUMPLIMIENTO
CON INDEPENDENCIA OPERATIVA.
HUMAN RIGHTS AS A GLOBAL PUBLIC GOOD:
ARTIFICIAL INTELLIGENCE AND MECHANISMS FOR
MONITORING COMPLIANCE WITH OPERATIONAL
INDEPENDENCE**

Alberto Delfín Arrufat Cárdua

Prof. Derecho Internacional Público

Universitat Jaume I de Castellón

Resumen:

La expansión de la inteligencia artificial (IA) está transformando profundamente el entorno digital en el que se desarrollan los niños, generando nuevas oportunidades para el ejercicio de sus derechos, pero también riesgos inéditos para su protección efectiva. Aunque los órganos de tratados de las Naciones Unidas, especialmente el Comité de los Derechos del Niño, han afirmado la plena aplicabilidad de los derechos humanos en los entornos digitales y desarrollado estándares interpretativos sobre privacidad, no discriminación, protección frente a la violencia y explotación de datos, dichos avances normativos no han ido acompañados de capacidades técnicas equivalentes para supervisar su cumplimiento en sistemas algorítmicos complejos. Este trabajo analiza el impacto de la IA sobre las funciones de monitoreo del subsistema convencional de protección de los derechos humanos de la ONU y sostiene que la ausencia de herramientas propias de inteligencia artificial, diseñadas conforme a estándares internacionales de derechos humanos, puede generar una situación de dependencia tecnológica indirecta respecto de arquitecturas algorítmicas externas orientadas por criterios ajenos a la lógica jurídica de protección. Particular atención se presta al riesgo de discriminación indirecta derivado del uso de variables aparentemente neutrales y a las dificultades de escrutinio derivadas de la opacidad de los modelos. Finalmente, se examinan iniciativas impulsadas en el seno de las Naciones Unidas para fortalecer capacidades institucionales en materia de gobernanza digital e inteligencia artificial, valorando su potencial contribución a la independencia operativa y eficacia futura de los mecanismos internacionales de supervisión.

Palabras Clave: derechos humanos, derecho del niño, inteligencia artificial.

Abstract: The rapid expansion of artificial intelligence (AI) is profoundly transforming the digital environment in which children develop, creating new opportunities for the enjoyment of their rights while also generating unprecedented risks to their effective protection. Although United Nations treaty bodies, particularly the Committee on the Rights of the Child, have affirmed the full applicability of human rights in digital environments and developed interpretative standards concerning privacy, non-discrimination, protection from violence, and

data exploitation, these normative advances have not been accompanied by equivalent technical capacities to monitor compliance within complex algorithmic systems. This article examines the impact of AI on the monitoring functions of the UN treaty-based human rights protection subsystem and argues that the absence of dedicated AI tools designed in accordance with international human rights standards may create a situation of indirect technological dependence on external algorithmic architectures driven by commercial, statistical, or functional objectives that differ from the legal rationale underlying human rights protection. Particular attention is paid to the risk of indirect discrimination arising from the use of apparently neutral variables and to the scrutiny challenges generated by algorithmic opacity. Finally, the article analyses recent initiatives adopted within the United Nations to strengthen institutional capacities in the fields of digital governance and artificial intelligence, assessing their potential contribution to the operational independence and future effectiveness of international human rights monitoring mechanisms.

Key Words: human rights, rights of the child, artificial intelligence

I.- Inteligencia Artificial como tecnología disruptiva.

La expansión acelerada de la inteligencia artificial (en adelante, IA) está transformando el entorno digital en el que se desarrollan los niños, generando oportunidades para el ejercicio de sus derechos¹, pero también riesgos significativos para su protección efectiva. La IA ha sido catalogada como la tecnología disruptiva más decisiva del S.XXI². Su capacidad para generar ventaja económica ha impulsado su presencia en diversidad de sectores como las redes sociales, los videojuegos, los asistentes virtuales, las plataformas educativas y los sistemas de recomendación que utilizan diariamente millones de niños³.

¹ Estrada señala -citando documentos de las Naciones Unidas- como posibles oportunidades para las niñas, la personalización de la educación y el acceso a nuevas habilidades tecnológicas (Estrada Tank, D.: Inteligencia artificial y derechos humanos de las mujeres y las niñas: sesgos, desafíos y oportunidades a la luz del Derecho Internacional. Cuadernos de Derecho Transnacional (marzo, 2026), Vol. 18, N.º 1, pp. 548-569, p. 557. ISSN 1989-4570 - www.uc3m.es/cdt - DOI: 10.20318/cdt.2026.10288).

² Para Becerra. y Magnani la IA es una de las tecnologías que se perciben sociológicamente como disruptivas (Becerra, G. y Magnani, E. (2024). La inteligencia artificial como tecnología disruptiva y sus sentidos sociales. *El Faro. Revista Digital de Docencia Universitaria*, 1(1), 44-59, p. 57); para Cernada Badía se trata de un giro copernicano no exento de riesgos que debe abordarse desde la regulación (Cernada Badía, Rosa. “Inteligencia artificial y Justicia en la Unión Europea: statu quo y perspectivas de futuro” Cátedra PAGODA, Universitat de València), 2025.

³ Para Cotino la inteligencia artificial se impulsa fundamentalmente por intereses económicos y competitivos (por eso) la ética de la IA ha de implementarse en los sectores empresariales, profesionales, tecnológicos e incluso en promotores de políticas públicas y de futuras regulaciones (Cotino Hueso, L. “Ética en el diseño para el desarrollo de una inteligencia artificial, robótica y big data confiables y su utilidad desde el Derecho,” *Revista Catalana de Dret Públic*, no. 58 (2019): 29-48, p. 44).

La IA se basa en el aprendizaje automático: un modelo aplica una función ajustando parámetros mediante la minimización de errores; a su vez, esas asociaciones se generalizan sobre nuevos datos sin nueva programación explícita de reglas. En otras palabras, que la IA no incorpora de forma automática principios de derechos humanos, ya que estos no emergen del proceso de optimización estadística del modelo, sino que dependen de decisiones humanas externas y de las asociaciones que va construyendo a partir del análisis de datos adquiridos.

En términos técnicos, su comportamiento está determinado por la función de pérdida, la distribución y curación de los datos de entrenamiento, y los mecanismos de evaluación y retroalimentación. Por lo tanto, que una IA tengan en cuenta los derechos humanos depende de su integración explícita: (i) mediante la formalización de restricciones normativas o penalizaciones en la función objetivo, (ii) a través del diseño y depuración del dataset para evitar sesgos o vulneraciones sistemáticas, y (iii) mediante procesos de evaluación, auditoría y ajuste del modelo -incluyendo feedback humano- que alineen la optimización de los parámetros *solo* con estándares de protección de derechos humanos.

II.- Subsistema de protección de los Derechos Humanos de la ONU y contexto digital.

El Comité de los Derechos del Niño (en adelante, el Comité) ha desempeñado un papel central en la interpretación evolutiva de la Convención sobre los Derechos del Niño (en adelante, CDN)⁴, al afirmar que los derechos en ella reconocidos son plenamente aplicables en entornos digitales y sistemas algorítmicos, lo que implica su extensión funcional a infraestructuras tecnológicas privadas y públicas que procesan datos de menores. En la Observación General número 25 de 2021 sobre los derechos de los niños en relación con el entorno digital, el Comité precisó que los Estados deben garantizar la protección frente a prácticas de perfilado, vigilancia, explotación comercial de datos y exposición a contenidos nocivos generados por procesos automatizados, en conexión con los artículos 16 -protección de la privacidad-, 17 -acceso a información- y 19 -protección frente a toda forma de violencia- de la Convención. Además, ha vinculado estas obligaciones con su artículo 36 que prohíbe otras formas de explotación -cuestión

⁴ Convención sobre los Derechos del Niño, adoptada el 20 de noviembre de 1989, 1577 UNTS. 3, entrada en vigor el 2 de septiembre de 1990.

relevante en contextos de monetización de datos infantiles y publicidad conductual⁵. Otros órganos de monitoreo de la ONU se han pronunciado en sentido similar: el Comité de Derechos Humanos ha establecido que la libertad de expresión se aplica plenamente a Internet y a los medios digitales, exigiendo que cualquier restricción cumpla los estándares de legalidad, necesidad y proporcionalidad⁶; también ha introducido consideraciones sobre el impacto de las tecnologías de vigilancia y sistemas digitales en el derecho a la vida⁷. Por su parte, el Comité para la Eliminación de la Discriminación contra la Mujer ha incorporado expresamente la dimensión digital de la violencia contra las mujeres, incluyendo el acoso en línea y la difusión no consentida de contenidos íntimos en relación con los artículos 1 y 2 de la Convención sobre la Eliminación de Todas las Formas de Discriminación contra la Mujer⁸. También el Comité de Derechos Económicos, Sociales y Culturales, en la Observación General número 24 de 2017 sobre el artículo 6 del Pacto Internacional de Derechos Económicos, Sociales y Culturales y en la Observación General número 25 (2020) sobre ciencia y derechos económicos, sociales y culturales, ha reconocido el impacto estructural de la IA en el ejercicio de derechos sociales, especialmente en materia de trabajo, educación y acceso a la información, destacando la necesidad de evitar que la brecha digital genere nuevas formas de desigualdad⁹. Estas líneas interpretativas se han reflejado igualmente en la práctica reciente de los exámenes periódicos a los Estados parte¹⁰.

Sin embargo, dicho desarrollo interpretativo no ha ido acompañado de la puesta a disposición de estos órganos de herramientas específicas de supervisión o verificación que permitan evaluar y comprobar de forma efectiva el cumplimiento de los derechos humanos en el entorno digital, especialmente en relación con el funcionamiento interno

⁵ CRC, *General Comment No. 25 (2021) on Children's Rights in Relation to the Digital Environment*, CRC/C/GC/25 (2 March 2021).

⁶ HRC, *Observación General núm. 34, Artículo 19: Libertad de opinión y libertad de expresión*, CCPR/C/GC/34 (12 de septiembre de 2011), párrs. 11–12 y 43.

⁷ HRC, *Observación General núm. 36 sobre el artículo 6 del Pacto Internacional de Derechos Civiles y Políticos, relativa al derecho a la vida*, CCPR/C/GC/36 (30 de octubre de 2018), párrs. 26 y 62.

⁸ CEDAW, *Recomendación General núm. 35 sobre la violencia por razón de género contra la mujer; por la que se actualiza la Recomendación General núm. 19*, CEDAW/C/GC/35 (26 de julio de 2017), párrs. 20 y 21.

⁹ CESCR, *Observación General núm. 24 (2017) sobre las obligaciones de los Estados en virtud del Pacto Internacional de Derechos Económicos, Sociales y Culturales en el contexto de las actividades empresariales*, E/C.12/GC/24 (10 de agosto de 2017), párrs. 13, 18 y 21; y CESCR, *Observación General núm. 25 (2020) sobre la ciencia y los derechos económicos, sociales y culturales (artículo 15.1.b del Pacto Internacional de Derechos Económicos, Sociales y Culturales)*, E/C.12/GC/25 (30 de abril de 2020), párrs. 13–16 y 21.

¹⁰ Observaciones finales del CEDAW sobre el noveno informe periódico de Alemania (CEDAW/C/DEU/CO/9) de 31 de mayo de 2023.

de la IA. Dicha ausencia es especialmente grave en un ecosistema tecnológico global caracterizado por la automatización, la opacidad de los sistemas y el tratamiento masivo de datos.

III.- Derechos humanos, control de cumplimiento e inteligencia artificial

Paradójicamente, en un momento en que la doctrina plantea la consideración de la dignidad de la persona como un valor autónomo de la sociedad internacional y su protección como *un bien de interés público global*¹¹, los órganos de monitoreo no sólo tienen graves problemas de financiación, sino que carecen de infraestructuras de IA propias y específicamente diseñadas conforme a los estándares interpretativos de derechos humanos. Por el contrario, gran parte de la información a partir de la cual realizan sus funciones de supervisión y recomendación se obtiene mediante canales tradicionales -visitas, reuniones, etc.- o mediante el análisis de información -que proviene de consultoras y organizaciones no gubernamentales- que pueden estar empleando modelos de IA desarrollados por terceros y optimizados a criterios comerciales, estadísticos o funcionales -sin perjuicio de las indicaciones (*prompts*) que pueda haberle facilitado el usuario-. Esto es preocupante en un ecosistema tecnológico global caracterizado por la automatización, la opacidad de los sistemas y el tratamiento masivo de datos.

Por eso, el subsistema de protección de los derechos humanos de origen convencional de la ONU puede experimentar en poco tiempo una situación de dependencia tecnológica indirecta respecto de arquitecturas algorítmicas externas cuya lógica de entrenamiento, criterios de optimización y mecanismos de corrección no necesariamente reflejan el enfoque de derechos humanos jurídicamente pretendido.

Uno de los retos más complejos que enfrentarán los Órganos de tratados es la valoración de la juridicidad de la actuación estatal ante la potenciación de la discriminación indirecta con el efecto de la IA. Por ejemplo, en aquellos casos donde el software utilice variables aparentemente neutras -como gustos musicales, acento, red social preferida, expresiones del lenguaje asociadas a salud mental- para predecir un resultado deseado. Si bien dichas variables no son en sí mismas características protegidas, están

¹¹ Cardona Llorens, J., *La protección internacional de la persona*. Valencia: Tirant lo Blanch, 2022, p. 28.

correlacionadas con el origen socioeconómico familiar o aspectos culturales del candidato. Su resultado es extrapolable a diversos ámbitos que afectan decisivamente a derechos de la infancia o a las personas de edad avanzada; por ejemplo: posibilidades de uso a término de sensores glucémicos dispensados por los sistemas nacionales de salud en función de asociaciones y correlaciones que permiten categorizan familias y hogares para predecir su grado de cuidado y atención al dispositivo.

Resulta probable que en el futuro los Estados sostengan que disponen de sistemas compatibles con las exigencias internacionales apoyándose en informes técnicos elaborados por terceros relativos a vigilancia educativa o asignación de prestaciones sociales, cuando en realidad la función objetivo ha sido optimizada exclusivamente conforme a criterios de eficiencia administrativa, reducción de costes o predicción estadística -sin incorporar restricciones relativas a no discriminación, interés superior del niño, privacidad ni proporcionalidad de manera adecuada-. También es previsible que la información remitida por los “amigos del Comité”¹² pierda progresivamente eficacia si no se apoya igualmente en sistemas de IA normativamente constreñidos a estándares de derechos humanos que no aseguren un buen escrutinio del Estado examinado.

La ONU ha empezado a trabajar en la dirección de fomentar la creación de herramientas propias de IA que cumplan con los derechos humanos y desligadas de aquellas otras generadas con fines distintos. En el plano estatal, la Declaración conjunta sobre inteligencia artificial y derechos del niño no se limita a exigir mecanismos de supervisión o mitigación de daños, sino que impone a los Estados deberes de capacitación de jueces y funcionarios, promoción de investigación y resto de la sociedad civil e inversión pública orientada específicamente a generar IA alineada con la efectividad de los derechos del niño¹³.

Por su parte, la Asamblea General de la ONU en el marco del fortalecimiento del

¹² En referencia al conjunto de actores externos que contribuyen al proceso de supervisión internacional mediante el envío de información especializada al Comité correspondiente. No equiparable a la figura procesal del *amicus curiae*, aunque comparte la finalidad de aportar elementos especializados destinados a mejorar la calidad deliberativa y el escrutinio internacional desarrollado por el Comité.

¹³ Declaración conjunta sobre inteligencia artificial y derechos del niño. Organización internacional de las Comunicaciones, Comité de los derechos del niño y otros Organismos del Sistema de la ONU. Noviembre 2025.

Sistema de Naciones Unidas, aprobó el 22 de septiembre de 2024 Resolución relativa al Pacto sobre el Futuro (Global Digital Compact) que reconoce expresamente la necesidad de “support capacity-building for developing countries to access, develop, use and govern digital technologies and artificial intelligence” y de fortalecer mecanismos multilaterales de cooperación y gobernanza digital dentro del sistema ONU. Un año más tarde se aprobó la Resolución que establece formalmente el “Independent International Scientific Panel on Artificial Intelligence” y el “Global Dialogue on Artificial Intelligence Governance”, institucionalizando capacidades permanentes de evaluación científica, coordinación y gobernanza internacional de la inteligencia artificial. Ambos instrumentos reflejan una evolución hacia el fortalecimiento de capacidades técnicas e institucionales propias en materia de IA por parte de la ONU, orientadas a garantizar una gobernanza digital compatible con el Derecho Internacional de los derechos humanos¹⁴.

Todavía es prematuro hablar de independencia operativa de los mecanismos de monitoreo de tratados de derechos humanos de la ONU en relación con herramientas de IA; sin embargo, su consecución es compatible con los caracteres del Sistema Internacional de protección de los derechos humanos porque: (a) pretende dotar a los mecanismos de herramientas que faciliten el funcionamiento de los Comités en su relación con los Estados y con las actividades que le son -directa o indirectamente- imputables; (b) tiene como objetivo la protección de la dignidad humana en un entorno -el digital- que incumbe al conjunto de la comunidad internacional en su conjunto; (c) se desarrollan para proteger el cumplimiento de obligaciones de naturaleza erga-omnes en su mayoría; que (e) serán utilizadas en mecanismos expresamente aceptados por los Estados; (d) ya sea para investigar vulneraciones a una persona -por ejemplo, la comunicación individual presentada por una víctima contra el estado por denegación de un trasplante basada en un algoritmo no entrenado contra la discriminación por edad- ya sea para investigar las vulneraciones -por ejemplo, por la proliferación de contenido discurso de odio en las redes- contra un colectivo determinado; y (e) que no conculca el principio de control subsidiario del sistema internacional respecto de los ordenamientos internos, dado que la aplicación de dichas herramientas operaría únicamente una vez

¹⁴ Resolución de la AG 325 relativa al mandato y modalidades para el establecimiento y funcionamiento del Panel Científico Internacional Independiente sobre Inteligencia Artificial y del Diálogo Mundial sobre la Gobernanza de la Inteligencia Artificial de 27 de agosto de 2025 (A/RES/79/325).

agotados los recursos internos.

IV.- Conclusiones

El monitoreo del cumplimiento de los derechos humanos es esencial para su aseguramiento. El Sistema universal de protección de los derechos de la persona ha ido progresivamente estableciendo mecanismos a tal fin. En las últimas dos décadas sus procedimientos se han perfeccionado resultando cada vez más incisivos.

La IA ha sido catalogada como la tecnología más disruptiva del SXXI; sus ventajas en coste económico han facilitado su extensión a multitud de sectores. En algunos de ellos -especialmente, allí donde la edad nos hace más vulnerables-, su aplicación plantea problemas de compatibilidad con los derechos humanos.

El Comité de los derechos del niño especialmente -y otros de forma menos explícita- han reaccionado recordando la plena vigencia de los derechos humanos en el entorno digital. No obstante, es evidente que la eclosión de la IA plantea retos importantes al modo en que desarrollan su función: las vulneraciones clásicas a la Convención ahora se van realizar de forma enrevesada; se van a potenciar nuevas formas de vulneración -especialmente, la discriminación indirecta-; conductas hasta ahora detectable mediante análisis documental exclusivamente humano van a dejar de serlo cuando la cantidad de información y la opacidad de los sistemas algorítmicos sean la base de la defensa de los estados en los exámenes periódicos ante la imputación de incumplimientos y la opacidad de los algoritmos -o su evaluación por terceros- la excusa para impedir su evaluación en investigaciones.

El presente trabajo ha planteado que la ausencia de herramientas de IA propias y normativamente constreñidas a estándares de derechos humanos puede dar lugar a una situación de dependencia tecnológica indirecta del subsistema de protección de los derechos humanos de la ONU respecto de arquitecturas algorítmicas externas que atienden objetivos distintos u obtienen su base de aprendizaje de fuentes ajenas al objetivo pretendido.

En los últimos años, se han adoptado en el seno de la ONU distintos acuerdos -en su

mayoría de naturaleza política o de soft law- que refuerzan la idea de fomentar la capacitación de funcionarios y el fomento de la inversión destinada a la generación de sistemas de IA desligados de modelos comerciales diseñados por las empresas del BigTech y enfocados al cumplimiento de los estándares de derechos humanos.

Si bien la utilización de dichas herramientas específicas e independientes por parte de los Comités con los fines previstos se prevé coherente con los objetivos del sistema universal de protección de los derechos humanos algunos aspectos de su uso pueden resultar controvertidos.