

IMPLICACIONES BIOÉTICAS Y PSICOPOLÍTICAS DE LA INTELIGENCIA ARTIFICIAL PARA LOS DERECHOS HUMANOS¹

BIOETHICAL AND PSYCHOPOLITICAL IMPLICATIONS OF ARTIFICIAL INTELLIGENCE FOR HUMAN RIGHTS

IMPLICAÇÕES BIOÉTICAS E PSICOPOLÍTICAS DA INTELIGÊNCIA ARTIFICIAL PARA OS DIREITOS HUMANOS

Hernando Barrios-Tao

(Universidad Militar Nueva Granada, Colombia)

hernando.barrios@unimilitar.edu.co

Recibido: 12/07/2024

Aprobado: 08/08/2024

RESUMEN

La inteligencia artificial invadió los ámbitos de la vida y se orienta a reconfigurar la vida misma, con impactos sobre los derechos humanos. Estas repercusiones sobre el desarrollo humano tienen, tanto detractores como defensores. El artículo se orienta establecer implicaciones bioéticas y psicopolíticas de la inteligencia artificial para los derechos humanos. El artículo se diseñó bajo el modelo PRISMA y recurrió a la hermenéutica analógica para interpretar los textos. Los resultados indican beneficios, amenazas y oportunidades para las libertades fundamentales y los derechos a la equidad, identidad, autodeterminación, autonomía, privacidad y bienestar. Las conclusiones refieren la integración de dispositivos, datos y algoritmos, que impactan los derechos humanos y plantean la necesidad de reflexionar sobre su recreación, reformulación o reconfiguración.

Palabras clave: inteligencia artificial. derechos humanos. bioética. psicopolítica. hermenéutica.

ABSTRACT

Artificial intelligence has invaded the areas of life and is aimed at reconfiguring life itself, with impacts on human rights. These repercussions on human development have both detractors and defenders. The article aims to establish bioethical and psychopolitical implications of artificial intelligence for human rights. The article, under the PRISMA model, uses analogical hermeneutics to interpret the texts. The results indicate benefits, threats and opportunities for fundamental freedoms and the rights to equity, identity, self-determination, autonomy, privacy, and well-being. The conclusions refer to the integration of devices, data, and algorithms, which impact human rights and raise the need to reflect on their recreation, reformulation, or reconfiguration.

¹ El artículo es producto del Proyecto de Investigación INV-HUM-3784 “Inteligencia artificial y derechos humanos de tercera generación”: Implicaciones educativas y psicopolíticas (2015-2021)”, financiado por la Vicerrectoría de Investigaciones de la Universidad Militar Nueva Granada, Bogotá – Colombia

Keywords: artificial intelligence. human rights. bioethics. psychopolitics. hermeneutics.

RESUMO

A inteligência artificial invadiu as áreas da vida e visa reconfigurar a própria vida, com impactos nos direitos humanos. Estas repercussões no desenvolvimento humano têm tanto detractores como defensores. O artigo tem como objetivo estabelecer implicações bioéticas e psicopolíticas da inteligência artificial para os direitos humanos. O artigo foi elaborado sob o modelo PRISMA e utilizou a hermenêutica analógica para interpretação dos textos. Os resultados indicam benefícios, ameaças e oportunidades para as liberdades fundamentais e os direitos à equidade, identidade, autodeterminação, autonomia, privacidade e bem-estar. As conclusões referem-se à integração de dispositivos, dados e algoritmos, que impactam os direitos humanos e suscitam a necessidade de refletir sobre a sua recriação, reformulação ou reconfiguração.

Palavras-chave: inteligência artificial. direitos humanos. bioética. psicopolítica. hermenêutica.

Introducción

La inteligencia artificial (IA) arribó a la integración de tecnologías, datos, algoritmos, realidades mixtas y aumentadas, con mayores posibilidades para invadir y remodelar los entornos sociales, culturales y políticos (Tegmark, 2017; O'neil, 2018). Asimismo, estos sistemas irrumpieron en la vida personal, relaciones, interioridad y conciencia, con capacidad para provocar cambios en neurotransmisores: serotonina, oxitocina, adrenalina y dopamina (Hernández, 2020).

Por otra parte, se avanzó de la interconexión a la interacción con los sistemas fuertes de IA, con posibilidades de reconfigurar la identidad y naturaleza humanas (Barrios-Tao, Díaz y Guerra, 2020) y hasta desnaturalizar la persona con mejoras transhumanistas que suplen supuestas debilidades de la condición humana (Castellanos, 2018). De manera particular, desarrollos de la bioingeniería plantean la dialéctica entre la biología convertida en tecnología y viceversa. En la primera relación dialéctica, los sistemas vivos (genes, células, órganos) se diseñan como dispositivos electrónicos, mientras que, en la segunda, los implantes electrónicos simulan sistemas y modifican seres vivos. De este modo, se crean artefactos inspirados en la integración de sistemas biológicos, cognitivos y socioculturales (Van Est & Gerritsen, 2017).

La problemática trasiega de riesgos en ámbitos sociales hacia peligros para los derechos, el desarrollo y la dignidad humana. En cuanto al desarrollo humano sostenible, la IA desafía la supervivencia planetaria, así como la armonía del ser humano con sus dimensiones de creatividad e innovación (UNESCO, 2015). Algunos avances neurotecnológicos amenazarían la dignidad humana con sus métodos para registrar, interpretar o alterar la actividad cerebral, que posibilitarían codificar, manipular pensamientos y alterar el ser humano (Jangwan, 2022), mientras que otros como las interfaces cerebro-computadora, permitirían una comunicación entre cerebro y mundo exterior, exportando datos o alterando la actividad cerebral (Yuste, Genser & Herrmann, 2021).

La Declaración mundial sobre Bioética y Derechos Humanos (2005) indicó principios para la relación entre derechos humanos, medicina y tecnologías, con el propósito de “promover el respeto de la dignidad y proteger los derechos humanos, velando por el respeto de la vida de los seres humanos y las libertades fundamentales” (UNESCO, 2005, art. 2). Sin embargo, dos décadas después se abre la necesidad de considerar los derechos bajo la óptica de los desarrollos neurocientíficos, los sistemas fuertes de IA y las tecnologías biomédicas invasivas (Ivanii, Kuchuk, & Orlova, 2020; Van Est & Gerritsen, 2017). Documentos globales plantean principios éticos para regular la IA y salvaguardar los derechos humanos: transparencia, justicia y equidad, no maleficencia, responsabilidad, privacidad, beneficencia, libertad y

autonomía, verdad, sostenibilidad, dignidad, solidaridad (Jobin, Ienca & Vayena, 2019). No obstante, se indica que las normas aún son vagas para ser aplicadas a estas complejas tecnologías (Barrio, 2021; Van Veen, 2018).

Por último, la relación entre IA y desarrollo humano se mueve entre detractores y defensores, entre quienes plantean visiones utópicas con beneficios y quienes indican riesgos y amenazas: bioconservadores que indican consecuencias desastrosas y trashumanistas con pretensiones de mejorar la vida (De Asís, 2020); “utópicos digitales” y “tecnoescépticos” (Tegmark, 2017). En las biotecnologías se mezclan dispositivos, datos y algoritmos, y se interpretan bajo los polos de utopías y distopías (James et al., 2023; Robles, 2020), promesas y desafíos (Bakiner, 2023), avances y peligros (Megías, 2022), impactos positivos y negativos (Van Est & Gerritsen, 2017), sueños y pesadillas (Martorell y Andonio, 2019), entusiasmos y temores (Pérez et al., 2017). Esta dialéctica también se consideró en la Declaración (UNESCO, 2005, art. 20) y se indicó la necesidad de “promover una evaluación y una gestión apropiadas de los riesgos” de las tecnologías conexas. Los nuevos avances trazan la necesidad de una interpretación equilibrada entre riesgos, beneficios y oportunidades. En este sentido, el artículo se orienta establecer implicaciones bioéticas y psicopolíticas de la IA para los derechos humanos.

Referentes teóricos: Inteligencias artificiales, bioética, psicopolítica y derechos humanos

El plural del constructo inteligencias artificiales (IAs) se determina por su polisemia, originada en la dinámica de sus desarrollos, los cuales arribaron a la integración de dispositivos, *big data*, técnicas, algoritmos, realidades mixtas y aumentadas. En cuanto a sus significados se pasó de la acción de diseñar dispositivos que funcionan como entidades, actúan, advierten e interactúan con su entorno (Robles, 2020), a considerar la IA como un saber multi e interdisciplinar que reúne matemáticas, informática, ingeniería, psicología, biología, estadística, entre otras (Stone, 2016; Zawacki et al., 2019, hasta la definición amplia e integradora de la Comisión Europea (2020, p.20):

Programas informáticos (y posiblemente también equipos informáticos) diseñados por seres humanos que, dado un objetivo complejo, actúan en la dimensión física o digital mediante la percepción de su entorno mediante la adquisición de datos, la interpretación de los datos estructurados o no estructurados, el razonamiento sobre el conocimiento o el tratamiento de la información, fruto de estos datos y la decisión de las mejores acciones que se llevarán a cabo para alcanzar el objetivo fijado.

Los dominios de actuación de las IAs se dinamizaron de los servicios (transportes, salud, seguridad, trabajo) (Stone, 2016), a los ámbitos educativos, jurídicos, políticos (Morán, 2021; Zawacki et al., 2019) y arribaron a la vida interior de las personas con fines comerciales (Han, 2014). Estas dinámicas permiten clasificarlas de acuerdo con la complejidad de sus acciones: de sistematizar datos y automatizar actividades, se pasó emular aspectos humanos. De este modo, se denominan IAs, estrechas, débiles, súper y fuertes (Baker & Smith, 2019), de acuerdo con sus capacidades para emular voz, percepción visual, lenguaje natural, cognición, emociones, inteligencia social (Kaplan & Haenlein, 2019) o imitar posibles dimensiones filosóficas como la conciencia, ética y moral (Colmenarejo, 2018).

En cuanto a la bioética y el cuidado de la vida, el ejercicio del poder pasó de la biopolítica, ejercida por el capitalismo sobre el cuerpo (Foucault, 2009; Galparsoro y Pérez, 2018; Cruz, 2017), a la consideración de los genes y el ADN como dispositivos biopolíticos (Sibilia, 2006), hasta la psicopolítica neoliberal que ejercita el poder sobre la interioridad (*psiche*), pensamientos, emociones y necesidades internas, mediante nuevas armas de gobernanza: motivación, ilusión, competencia, originalidad, iniciativa (Han, 2014). Estas formas de manipulación generan adicciones y marcan un estadio sofisticado del ejercicio del poder, la “psicopolítica digital”: un nuevo y más efectivo control ejercido desde dentro (Espinosa, 2018; Cruz, 2017) y capacitado para leer y controlar pensamientos (Han, 2014). Este poder sobre el ser humano, centro de la bioética, se realiza mediante datos configurados por algoritmos, códigos de *software* que procesan instrucciones en cadena, con el fin de obtener determinados resultados (Barrios-Tao y Díaz, 2024; Hill, 2016). Estos códigos o unidades epistémicas integran elementos cognitivos y valores éticos con propósitos mercantiles (O’neil, 2018; Cardon, 2018; Comisión Europea, 2018; Arbulo, Olveira y Wilkins, 2018).

La integración de IAs fuertes y psicopolítica conduce la mirada hacia los derechos humanos. Mientras que unos consideran necesario su ampliación, actualización, reformulación o reconceptualización (Razmetaeva, Barabash, & Lukianov, 2022; Morales, 2018), otros plantean una nueva generación de derechos asociados con las innovaciones de las IAs (Castelblanco, 2021; Perepolkin, Perepolkin, & Averianova, 2021; Ivani, Kuchuk, & Orlova, 2020; Arbulo, Oliveira y Wilkins, 2018; Aguirre y Manasía, 2015). Esta nueva generación configuraría derechos a la identidad, la construcción individual y la protección de la dignidad, libertad, autonomía y privacidad (De la Quadra-Salcedo, 2018; Pérez, 2018), así como a la inclusión digital, a la información y conocimiento (Cova-Fernández, 2022). El objetivo de esta nueva generación sería “corregir las externalidades negativas y los déficits de justicia e igualdad que están propiciando la falta de regulación jurídica adecuada que acompaña la revolución digital” (Barrio, 2021, p. 207). Sin embargo, autores como Torres (2018) desconocen estas clasificaciones y recurren a las escuelas de derecho (naturalista, deliberativa, protesta, discursiva) para comprenderlos de forma más holística.

Otros autores retoman el planteamiento sobre el neuroderecho (Taylor, Harp & Elliot, 1991) y proponen nuevos derechos: libertad cognitiva, privacidad e integridad mental, continuidad psicológica (Ienca & Andorno, 2017), así como la defensa y cuidado del desarrollo humano: identidad personal, libre albedrío, entre otros (Yuste, Genser & Herrmann, 2021; Borbón, Borbón y Laverde, 2021; De Asís, 2020). Por último, para otros la biotecnología implicaría el surgimiento de derechos somáticos asociados con el cuerpo humano y estructurados en grupos: eutanasia; reproductivos y disposición de órganos y tejidos; esfera y cambios sexuales, entre otros (Perepolkin, Perepolkin, & Averianova, 2021; Shevchuk et al., 2021; Ivani, Kuchuk, & Orlova, 2020).

Metodología

La revisión de literatura se orientó por la pregunta acerca de las implicaciones bioéticas y psicopolíticas de los sistemas de IA para los derechos humanos y siguió los pasos declarados por el modelo PRISMA (*Preferred Reporting Items for Systematic Review and Meta Analyses*) (Page et al., 2021), en sus cuatro fases: identificación, filtrado (o cribado), elegibilidad e inclusión (González et al., 2011). La identificación inicial de los textos (N=184) se realizó mediante la combinación de palabras clave en castellano y sus correspondientes en inglés y portugués: “inteligencia artificial”, “derechos humanos”, “bioética”, “psicopolítica”, combinando operadores booleanos, de acuerdo con las interfaces de *Science Direct*, *Scopus*, *EBSCO*, *Scielo*, bajo el marco temporal (2015-2023). Posteriormente se procedió al filtrado (excluidos n=57; selección temporal N=127), mediante filtros sobre tipología de documentos (artículos de investigación, tesis, documentos de instituciones internacionales). La tercera fase de elegibilidad consideró factores de accesibilidad y calidad de las fuentes (excluidos n=44; selección temporal N=83). Por último, se procedió a la inclusión de los textos pertinentes para responder la pregunta orientadora y se consideraron algunos textos de autores relevantes fuera del marco temporal (excluidos n=15; inclusión final N=68).

Para interpretar los textos incluidos se recurrió a la hermenéutica analógica determinada como una teoría aplicable a problemáticas sociales y a fenómenos que se mueven entre extremos interpretativos. Para Beuchot (2016, p.114) esta hermenéutica se puede aplicar como “política de la interpretación” a la vez que “interpretación de la política” y podría contribuir en “des-totalizar los significados unívocos y en re-totalizar los significados equívocos”, para llegar a una prudencia individual y colectiva orientada a la “interpretación del hombre, del ser humano, para comprenderlo y, a partir de eso, procurarles una ética, un derecho y una política que le sean adecuados, que le resulten convenientes”. En el caso de las implicaciones de las IAs para los derechos humanos, la interpretación se mueve entre beneficios para fortalecer principios bioéticos como la justicia y la autonomía, y las amenazas del poder sobre libertades fundamentales, privacidad o bienestar, entre otros. La hermenéutica analógica estructura un equilibrio entre la univocidad y la equivocidad, cuando se consideran implicaciones en términos de riesgos, beneficios y oportunidades.

El diseño de la interpretación siguió las tres fases que estructuran y orientan la hermenéutica analógica: sintáctica, semántica y pragmática (Beuchot, 2016). En la fase sintáctica se organizaron los textos en una matriz diseñada, a través de las unidades temáticas emergentes orientada por la pregunta investigativa: “libertades fundamentales” (f), “equidad” (e), “identidad” (i), “autonomía” (a), “privacidad” (p), “bienestar” (b). La fase semántica identificó y caracterizó la univocidad (U) o visiones utópicas, así como la equivocidad o visiones distópicas (D). Por último, mediante una matriz de interpretación se procedió a la fase pragmática para establecer la dimensión analógica (A) o *phronética*. De este modo, los resultados de la interpretación se presentan de forma narrativa siguiendo cada una de las seis (1-6) temáticas emergentes (f, e, i, a, p, b) con sus respectivas voces unívocas (U), equívocas (D) y analógicas (A).

Resultados: Libertades fundamentales (f-A): entre soberanos (U), manipulados y gobernados (D)

La Declaración señala el deber de “respetar plenamente la dignidad humana, los derechos humanos y las libertades fundamentales” (UNESCO, 2005, art. 3). Las primeras implicaciones se orientan a estas libertades: de opinión, expresión, pensamiento, consciencia, entre otras. Algunas voces utópicas (U) indican beneficios para estas libertades por la posibilidad de expresarse libremente y deambular sin límites en internet (Castelbanco, 2021). Por otra parte, más información circulante en la red, posibilidades de personalizar contenidos y aumento de canales de comunicación, contribuirían con las libertades de pensamiento, opinión y expresión (Larrondo y Grandi, 2021; Van Est & Gerritsen, 2017; Ienca & Andorno, 2017).

Las visiones distópicas (D) plantean riesgos de manipulación y restricción para la libre movilidad y la seguridad debido a la capacidad de los sistemas para proporcionar detalles sobre movimientos individuales y predecir ubicaciones futuras (Andersen, 2021). Un segundo ámbito amenazado es la libre expresión, cuando en los sistemas de IA se desbordan los límites, llegando al irrespeto, desprecio y violencia (Larrondo y Grandi, 2021) o cuando se extralimita la capacidad de pronunciarse hasta denigrar del otro; este riesgo aumenta debido a la lentitud para desarrollar jurisprudencia y normas consistentes para sancionar estas agresiones (Castelblanco, 2021; Morales, 2018; Van Est & Gerritsen, 2017) y a las condiciones de inmaterialidad e invisibilidad para el escrutinio público de estas acciones en la red (Arbulo, Oliveira y Wilkins, 2018). En esta violencia digital se suplantán usuarios y se configuran respuestas automáticas con cuentas perfiladas, mediante *bots* emuladores del lenguaje natural (Andersen, 2021).

El tercer ámbito amenazado es la libertad mental. Ienca y Andorno (2017) refieren investigaciones sobre el control cerebral, no invasivo, desarrollado en dispositivos de comunicación mediante *neurogadgets* y la lectura de ondas cerebrales. El riesgo para la libertad mental y el autogobierno radica en el potencial de la ingeniería neuronal y de imágenes cerebrales para acceder a la información mental y derrumbar la otrora fortaleza inexpugnable de la mente. Asimismo, Borbón, Borbón y Laverde (2021) indican técnicas como la optogenética orientadas a modificar genéticamente las neuronas generando su fotosensibilidad, mediante virus, con la posibilidad de excitarlas o inhibirlas con la luz. Por otra parte, la estimulación cerebral con ultrasonidos posibilitaría cierto control sobre el comportamiento (Kubaneck et al., 2020).

Otro ámbito en riesgo es la libertad individual debido a la manipulación y gobernanza en el novedoso panóptico digital (Robles, 2020; Han, 2014) posibilitadas por la pérdida del control sobre datos de la vida privada, suministrados a los sistemas de IA. Allí, nombre, fama, imagen y perfil, se utilizan para fines no contemplados por los usuarios (Castelblanco, 2021). Las configuraciones algorítmicas perfeccionan métodos para controlar y vigilar, a través de la recopilación, configuración y manipulación de esta información personal (Bakiner, 2023; Barrio, 2021; Robles, 2020; Cotino, 2019; Arbulo, Oliveira y Wilkins, 2018). En síntesis, se podría configurar un “autoritarismo digital” para controlar las actividades ciudadanas con prácticas de censura y restricción de las libertades fundamentales (Castellanos, 2019, p.7).

En cuanto a la gobernanza, Guerrero (2023) indica mecanismos de dominación y limitación de los derechos como los procesos de reeducación a los usuarios al convertirlos en consumidores y colonizar sus conciencias, mediante la configuración algorítmica de ciertos valores. Por otra parte, la libertad para recibir o rechazar información se coarta a través de estrategias como las cámaras de eco y los filtros burbuja, los cuales determinan la información que se debe mostrar y nublan el desarrollo de la libre opinión, la libre expresión y la garantía de recibir información libre y sin interferencias (Megías, 2022; Larrondo y Grandi, 2021; Nguyen, 2020; Pariser, 2017). Otra estrategia es el *Big Nudging* mediante el cual no se indica directamente el modo de obrar, sino que se utiliza una “macro-insinuación” para redireccionar el comportamiento individual (Castellanos, 2019). Por último, otras investigaciones evidencian prácticas de gobernanza como la bio y la bibliometría, orientadas a medir la supuesta calidad de los individuos (Feldman & Sandoval, 2018; Lupton, 2016) y a monitorear comportamientos individuales, encaminados a determinar la calidad institucional (Lim, 2019; Saura y Bolívar, 2019).

Cada una de las analogías (A) resaltarán y exhortarán al cuidado a la dignidad humana y se interpretará por el equilibrio entre la ampliación, recreación o reformulación de los derechos humanos, con relación a los avances de las IAs. Esta primera visión *phronética* (A) inicia con un llamado a las neurotecnologías para preservar la “libertad mental”, mediante el cuidado de la capacidad personal para ejercer el autogobierno sobre pensamientos y comportamientos, con el fin de preservar la mente, como refugio de la libertad personal (Ienca & Andorno, 2017). Este derecho ampliado de “libertad mental” se relaciona con el planteamiento de De Asís (2020) quien retoma la propuesta de Yuste, Genser y Herrmann (2021) acerca de su iniciativa para impulsar el “neuroderecho a la libertad cognitiva”, determinado como la decisión de las personas sobre su deseo o no de utilizar o rechazar las neurotecnologías en su dimensión corporal.

Por otra parte, algunos textos trazan deberes para los configuradores de sistemas de IA. Cualquier configuración debe preservar el derecho a no ser medido, analizado o entrenado (*coached*). En este sentido, la libertad para recibir información no se debería promover, ni obstaculizar, mediante estrategias que orienten, manipulen y gobiernen a los usuarios (Cotino, 2019; Van Est & Gerritsen, 2017). Por último, se debería fortalecer y garantizar la moderación en la generación de contenidos, con la posibilidad de ser revisados, supervisados y removidos por las personas y así garantizar sus libertades fundamentales (Larrondo y Grandi, 2021).

Equidad (e-A): entre incluidos (U) y marginados (D)

La equidad, principio fundamental de la bioética, se plantea a partir del respeto por la “igualdad fundamental de todos los seres humanos en dignidad y derechos” (UNESCO, 2005, art. 10). La Declaración señala que “ningún individuo o grupo debería ser sometido por ningún motivo” a fenómenos de “discriminación o estigmatización alguna” (art. 11).

Las visiones utópicas (U) se sustentan sobre el razonamiento de que, a más desarrollos de IA, mayor inclusión y equidad. En este sentido, *internet* se determinaría como una estructura social participativa, que posibilitaría democratizar el conocimiento, empoderar personas y fortalecer procesos de inclusión social (Guerrero, 2023). Una inclusión fortalecida mediante la alfabetización digital que integraría comunidades a través de avances tecnológicos, favoreciendo la participación democrática (Aguirre y Manasía, 2015).

En contraste, las distopías (D) refieren la exclusión desde diferentes ópticas. En primer lugar, la desigualdad entre una minoría con posibilidades, frente a una mayoría que no puede conectarse a la red, ni acceder a dispositivos digitales (Castelblanco, 2021). Esta brecha digital margina ahora por la limitación de acceso a condiciones técnicas, económicas y culturales, para interactuar y participar en las decisiones políticas sobre el desarrollo social (Guerrero, 2023; Barrio, 2021; De Asís, 2020). Una segunda óptica, contextualizada también en las desigualdades sociales, limitaría la posibilidad de acceder a dispositivos, desarrollar habilidades digitales (Castellanos, 2018) y acceder a diagnósticos o tratamientos en el ámbito de la telemedicina (Ishengoma, 2022).

En otro ángulo se plantea la vulneración de los derechos en términos de discriminación. Por una parte, se excluyen personas a causa de los valores e intereses de quienes configuran los algoritmos (O’neil, 2021; Van Est & Gerritsen, 2017) para designar recursos para la educación, la atención sanitaria o los servicios públicos de ciudades inteligentes (Lane, 2023; Human Rights Council, 2021). Algunos sistemas con respuestas automatizadas discriminan e impiden acceder a servicios, mediante filtros configurados con criterios asociados a factores culturales, económicos, étnicos o religiosos (Bakiner, 2023; Castelblanco, 2021; O’neil, 2021). Andersen (2021) presenta estrategias nocivas de configuraciones algorítmicas: perfiles discriminatorios mediante reconocimiento facial; selección, segregación y vigilancia masiva; propaganda selectiva, enlaces o fuentes falsas; perpetuidad de prejuicios jurídicos mediante calificación de riesgo y reincidencia; perfilamiento sesgado por raza o condición social.

En el ámbito biojurídico, el riesgo discriminatorio es para la justicia debido a la limitación de la mayoría para acceder a procedimientos judiciales en escenarios digitales, así como para el derecho a un juicio justo debido a que la jurisprudencia no avanza a la par de los avances tecnológicos y no se reconfiguran, ni se plantean nuevos derechos para afrontar las neurotecnologías (Van Est & Gerritsen, 2017). En la esfera sanitaria, Andersen (2021) indica que configuraciones algorítmicas podrían recomendar diferentes tratamientos, en función del seguro del paciente, su capacidad de pago y negar atenciones vitales por causas socioeconómicas, discriminando y perjudicando grupos marginados. Asimismo, se podrían manipular decisiones médicas con recomendaciones sobre el trato a pacientes perfilados: posibilidad de discriminación genética, cuando se tienen visiones exclusivamente biológicas de las personas (Faralli, 2015); atención sanitaria inequitativa por configuraciones sesgadas que excluyen pacientes con enfermedades raras o desfavorecidas en ensayos clínicos (Ishengoma, 2022). En síntesis, Van Veen (2018) plantea la inequidad en términos globales: perturbación grave de la distribución de poder en el mundo al crear nuevas formas de opresión y afectar a los más vulnerables.

En términos de repensar los derechos, la visión fronética (A) se orienta a considerar las oportunidades de las IAs para una mayor circulación y mejor aprovechamiento de los avances científicos, en beneficio de todos los ámbitos sociales y con el acceso equitativo a estos desarrollos (Toboso y Aparicio, 2019; Villalba, 2020). Van Veen (2018) plantea el hecho de que los mismos derechos humanos como lenguaje y marco legal, son una fuente de poder debido a su legitimidad moral. En consecuencia, se deberían cuidar para evitar la probabilidad de utilizarlos como poder para causar daño. Por último, es necesario fortalecer prácticas de inclusión en el ámbito digital para interactuar mediante redes de información y acceder a nuevos servicios, con el fortalecimiento de políticas de educación ciudadana orientadas a crear una inteligencia colectiva para afrontar un mundo globalizado (Comisión Europea, 2018; Baker & Smith, 2019).

Identidad y autodeterminación (i-A): entre subjetividades (U) y perfiles creados (D)

La Declaración aborda el derecho a la identidad, en su momento vinculado con la relación tecnologías, individuos y sociedad, y plantea una visión integral que comprenda sus dimensiones biológicas, psicológicas, sociales, culturales y espirituales (UNESCO, 2005, prefacio).

Las voces utópicas (U) consideran que más acceso a la información empoderaría la identidad y la autodeterminación individual. En cuanto a la identidad, las tecnologías posibilitan nuevas expresiones relacionales y configuran un yo más abierto y flexible, así como un yo múltiple con mayores posibilidades de lúdica y auto-creatividad (Illou, 2007). Por su parte, Floridi et al. (2018, p.692) indican que las IAs podrían “estar al servicio de la autodeterminación humana y fomentar la cohesión social”, sin peligros para la dignidad y el florecimiento de lo humano.

Los riesgos (D) se plantean con base en el potencial de la IA para crear nuevos modos de identidad; ciber-identidades heterónomas configuradas a partir de dispositivos mediáticos y no de valores individuales. Estas identidades adquieren diversas denominaciones: “homo zappiens”, “tecnopersona”,

“zoon elektronikón” (Echeverría, 2016; Huici y Davila, 2016). En todos los casos, las identidades se producen, desarrollan y determinan a través de las interacciones y relaciones mediadas por la IA.

Algunos textos refieren el uso de las corrientes positivas y negativas emocionales como estrategia para alterar, tanto la identidad como la autodeterminación individual. Las emociones, convertidas en inscripciones digitales (Ghotbi, 2023; Cardon, 2018), se manipulan mediante los datos emocionales acopiados, con los cuales se configuran perfiles o identidades falsas, se ejerce el poder y se subyuga al individuo asombrado y fascinado (Piñar, 2018). En consecuencia, los riesgos se intensifican por la posibilidad de construir y deconstruir perfiles, con o sin el consentimiento personal, mediante datos reales o ficticios. Esta configuración de perfiles con estadísticas manipuladas y configuradas por otros amenaza los ámbitos sociales, sanitarios y laborales (Beever & Taylor, 2022; De la Quadra-Salcedo, 2018).

A partir de la identidad, comprendida como el factor individual que permite experimentarse a sí mismo como la misma persona persistente en el tiempo, Ienca y Andorno (2017) plantean la amenaza de alterarla y desajustar personas, mediante la configuración de perfiles y avatares, así como por el desdoblamiento del individuo en la dialéctica persona-personaje. El personaje se vincula con la reputación *on-line* que termina devorando a la persona, condicionando las posibilidades de la vida real y cosificando la identidad, con el resultado de “una vida sin yo” (Suárez, 2014, p.39) o una identidad social distorsionada (Faralli, 2015). Por otra parte, cuando dispositivos neuronales se usan para estimular o modular la función cerebral podrían alterar los estados mentales críticos para la personalidad y afectar la identidad, es decir, “la percepción de las personas sobre su propia identidad” (Ienca & Andorno, 2017). Algunas técnicas de ingeniería de la memoria pueden afectar la identidad al eliminar, alterar, agregar o reemplazar selectivamente recuerdos relevantes para su autorreconocimiento como personas (Jangwan, 2022). En cuanto a la autodeterminación, Ienca y Andorno (2017) reportan casos de aplicación de neurotecnologías con efectos sobre el comportamiento: aumento de impulsividad y agresividad; cambios en el comportamiento sexual, la personalidad y el conocimiento de sí. Asimismo, con los avances de la terapia génica se podría influir en la transmisión del patrimonio genético a las generaciones sucesivas (Faralli, 2015).

La visión *phronética* (A) inicia con el llamado a plantear acciones frente al desafío de la modelación de la identidad y la conciencia mediante IAs (Guerrero, 2023). Los planteamientos de Valdés (2015, p.1220) indican la necesidad de cuidar los derechos orientados a la protección de la “identidad genética”, debido al riesgo de afectación personal en su “composición ontológica por los avances y aplicaciones de la biomedicina y la biotecnología”. Al mismo tiempo que los avances neurotecnológicos deberían florecer neuroderechos protectores de los riesgos. Es el caso del “derecho a la continuidad psicológica” que cuida de la preservación de la identidad personal y la prolongación de los pensamientos, preferencias y elecciones de una persona. Asimismo, el derecho a la integridad mental, que podría salvaguardar tanto la mente como sus pensamientos, creencias y convicciones de las posibilidades de restricciones procedentes de factores externos, así como de las amenazas a la mente como refugio también de la identidad y la autodeterminación individual (Ienca & Andorno, 2017). Por último, el derecho a heredar un patrimonio genético no modificado (Faralli, 2015).

Autonomía y responsabilidad (a-A): entre independientes (U) y heterónomos (D)

El principio bioético de autonomía es fundamental para la Declaración: “Se habrá de respetar la autonomía de la persona en lo que se refiere a la facultad de adoptar decisiones, asumiendo la responsabilidad de éstas y respetando la autonomía de los demás” (UNESCO, 2005, art. 5). Tanto el libre albedrío como la autonomía forman parte del desarrollo humano y posibilitan asumir responsabilidades individuales (Yuste, Genser & Herrmann, 2021).

Voces utópicas (U) relacionadas con el fortalecimiento de la autonomía y la responsabilidad se fundamentan en la contribución de las IAs para disponer de más información e interacción social, así como para mejorar capacidades y eficiencia personales (Verdegay et al., 2021; Villalba, 2020), como el

caso de las habilidades de razonamiento clínico involucradas en el diagnóstico y detección de enfermedades (James et al., 2023). Asimismo, cuando se participa de forma libre en actividades sociales, intelectuales y culturales, configuradas en los sistemas de IA (Floridi et al., 2018, Tegmark, 2017). Por último, las neurotecnologías al mejorar aspectos cognitivos como la memoria, el conocimiento, el razonamiento y la toma de decisiones contribuirían con mejorar la autonomía (Jangwan et al., 2022; Borbón, Borbón y Laverde, 2021).

En contraste, los riesgos (**D**) inician cuando se entregan datos privados, insumo para entregar la autonomía personal y actuar de forma heterónoma: usuarios seducidos y movilizados por los bombardeos de los configuradores de IA, quienes se valen de gustos, preferencias y pensamientos suministrados al producir y consumir información. Esta gestión algorítmica de gustos, recuerdos y deseos influye en la toma de decisiones individuales, en detrimento, tanto de la autonomía y capacidad de decisión, como de la participación ciudadana, debido a una configuración obediente a cálculos y acomodada a situaciones, es decir, una “subordinación a los datos” (Castellanos, 2018, p.127; 2019). De este modo, la inmersión y absorción en dispositivos atrofian el desarrollo de los usuarios como seres independientes (Morán, 2021; Van Est & Gerritsen, 2017). En el ámbito médico, Ishengoma (2022, p.97) refiere el dilema denominado la “mente médica colectiva” que aborda la transferencia de la autoridad de los médicos humanos a los algoritmos, cuando se pasa de utilizar las IAs como apoyo a las decisiones, a convertirlas en centrales de toma de decisiones médicas, con el riesgo de soslayar principios bioéticos como la beneficencia y no maledicencia. En este ámbito, James et al. (2023) plantean la amenaza de descalificar los roles y habilidades del personal médico.

Por otra parte, cuando no existe un consentimiento digital sobre el uso de los datos generados y acopiados se arriesga la responsabilidad. El riesgo radica, tanto en la perpetuidad de quienes los almacenan, comercializan y reconfiguran permanentemente, como en la pérdida del control personal, convirtiéndose en una mercancía circulante a perpetuidad (Hernández, 2020). Asimismo, se puede invadir y difuminar la responsabilidad individual, debido a que la celeridad de los avances y los múltiples actores que intervienen la configuración de los sistemas amenazan la capacidad de la industria tecnológica para crear protocolos y herramientas, que regulen y determinen responsabilidades de quienes intervienen en la cadena de producción, configuración, gestión y mercadeo (Castelblanco, 2021). En este sentido, la responsabilidad de diseñadores, operadores, usuarios y de otras personas no visibles en estos ecosistemas, se puede difuminar o esfumar (Villalba, 2020; Morales, 2018).

Unos autores refieren estrategias conducentes a distorsionar la toma de decisiones minando la autonomía. La inmersión en burbujas ideológicas configuradas con valores y contenidos, encierra a las personas en sus gustos y elecciones, focalizándolos en sus probabilidades, con la anulación de su autonomía y la posibilidad de reescribirse, recrearse y proyectarse (Nguyen, 2020; Cardon, 2018). Por su parte, las cámaras de eco, configuradas con datos privados entregados por los sujetos, resuenan y repiten la desacreditación de voces, conceptos disonantes y disconformes con las posiciones individuales. Los resultados se evidencian en la homogeneización de individuos, comunidades y sociedad (Nguyen, 2020; Comisión Europea, 2018).

Por último, otros riesgos provienen de avances neurotecnológicos capacitados para estimular el cerebro e influir, tanto en su comportamiento, como en sus recuerdos (Yuste, Genser & Herrmann, 2021). En este sentido, se podría impactar la denominada “computación neuronal” con la posibilidad de realizar un “hacking cerebral malicioso”, análogo al realizado en computadoras y datos digitales, y manipular directamente la información neuronal de una persona (Ienca & Andorno, 2017). Por último, la posibilidad de mejorar rasgos cognitivos podría alterar dimensiones humanas, algunas veces motivadas por satisfacción de caprichos individuales, en contra del ordenamiento social (Borbón, Borbón y Laverde, 2021).

La visión *phronética* (**A**) también se plantea como llamada para salvaguardar la autonomía y la responsabilidad en todos los intervinientes en las cadenas de producción, configuración y uso de los sistemas de IAs (Nunes, Miranda y Dadalto, 2022; Villalba, 2020), así como la consideración de crear

un neuroderecho orientado al cuidado de la integridad psíquica (Ienca & Andorno, 2017). Más allá de los consentimientos libres, expresos e informados (UNESCO, 2005) se deberían reconfigurar y rediseñar estos consentimientos frente a los lenguajes, funcionamientos y actuaciones de los avances de las IAs. Por último, Verdegay et al. (2021) plantean recomendaciones sociales y educativas sobre la claridad para los usuarios acerca del uso de la información, la preservación de su autonomía, sin el abuso de sus perfiles personales y el fomento de la formación integral sobre aspectos prácticos y potencialidades de las IAs.

Privacidad (p-A): entre la confidencialidad (U) y la intimidad descubierta (D)

El alcance del respeto a la privacidad se relaciona con “la confidencialidad de la información” y se indica que “no debería utilizarse o revelarse para fines distintos de los que determinaron su acopio o para los que se obtuvo el consentimiento” (UNESCO, 2005, art. 9). Sin embargo, este alcance se reconfigura por los avances de las IAs que arribaron a la interioridad individual. Voces utópicas (U) se mueven entre la esfera de confidencialidad promovida por los sistemas de seguridad de estos avances tecnológicos (Megías, 2022; Floridi et al., 2018) y la posibilidad de la IA para identificar ciberataques y salvaguardar sistemas de información sensibles como los del ámbito sanitario (Ishengoma, 2022).

El riesgo para la privacidad (D) continúa en la esfera informativa, no obstante, ahora se relaciona con datos privados entregados en dispositivos digitales (Megías, 2022; Castelblanco, 2021), cuya recopilación se usa para configurar perfiles y comercializarlos, lo cual amenaza el derecho a la privacidad y otros conexos (Bakiner, 2023 ; Human Rights Council, 2021). La posibilidad de configurar ciudades inteligentes con esta información aumenta los riesgos para la privacidad (Lane, 2023; Donahoe & Metzger, 2019). Más aún, la difuminación de los límites entre espacios físicos y digitales, entre información y datos privados, aumenta la amenaza debido a que también “las fronteras de los conceptos ‘privacidad’ y ‘libertad’ se difuminaron” (Arbulo, Olveira y Wilkins, 2018: 8).

Por otra parte, la información ubicada en formatos físicos era visible y se podía cancelar una vez alcanzase su objetivo. En contraste, en la información privada mediante la cual se configuran sistemas de IA, se cancelan la autoría del individuo y su dominio absoluto. Esos datos pueden ser modificados, multiplicados, comunicados y asumen un carácter “cuasi eterno” (Rallo, 2010, p.105) con el riesgo de afectar la persona *ad eternum*, cuantas veces alguien los obtenga: se podrían perjudicar personas si se convierten en instrumento de difamación y agravio contra la intimidad al exponerlos públicamente (Suárez, 2014). En el ámbito sanitario, la digitalización de historias clínicas creó unos nichos de datos biológicos, registros y bioinformática, los cuales se almacenan, comparten y distribuyen con amenazas para la privacidad y confidencialidad personal (Klavetter et al., 2023; Ivanson et al., 2023; Beever & Taylor, 2022; Nunes, Miranda y Dadalto, 2022; Karabekmez, et al., 2021).

El riesgo para la privacidad arriba a la *psique*, es decir, avanza de la privacidad de datos, a la del cerebro y la mente. El potencial de acceder a información almacenada en el cerebro amenaza lo más íntimo de la persona, la mente: “la presunción de privacidad mental ya no es una certeza” (Yuste, Genser & Herrmann, 2021, p.159). Más aún, la posibilidad de acceder a las ondas cerebrales y a la interioridad personal podría desembocar en daños psicológicos y mentales. No se desarrollan aún herramientas o dispositivos para filtrar la información privada circulante en las ondas cerebrales y así evitar riesgos cuando se acceda de forma ilegítima a esa información (Shevchuk et al., 2021; Ienca & Andorno, 2017). En síntesis, el riesgo para la privacidad mental está conexo con la posibilidad detectar, mapear e interpretar la actividad cerebral (Jangwan et al., 2022).

La visión analógica (A) se orienta a plantear derechos que cuiden el equilibrio entre la naturaleza abierta y atemporal de los sistemas digitales y la protección de los datos privados y la mente de las personas (Rallo, 2010; Tegmark, 2017). De ahí, la ampliación del derecho a la privacidad mental para proteger la información cerebral registrada en neurodispositivos, que se puede compartir en el ecosistema digital (Ienca & Andorno, 2017). Además, preservar la intimidad mediante un derecho al olvido de los datos difundidos en la red (Barrio, 2021; Morales, 2018; Suárez, 2014). Por otra parte, el respeto a la vida

privada se debería recrear en relación con el derecho al respeto a la vida familiar, a establecer y a desarrollar relaciones interpersonales afectivas, sin interferencias ni invasiones a la privacidad, ni a los principios y valores (Van Est & Gerritsen, 2017). Asimismo, las personas deberían considerarse como tales y no como una “mera suma de datos informáticos” (Faralli, 2015, p.20) y preservar su intimidad como derecho a permanecer en paz “y a no ser molestado en ese castillo inexpugnable que supone la propia vida” (Suárez, 2014, p.36).

Bienestar (b-A): entre calidad de vida (U) y riesgos para la salud integral (D)

El bienestar abarca la salud, el desarrollo integral y “debería tener prioridad con respecto al interés exclusivo de la ciencia o la sociedad” (UNESCO, 2005, art. 3). Las visiones utópicas (U) se cimentan en el bienestar ofrecido por la IA en los ámbitos de actuación del ser humano: transportes, salud, educación, seguridad, entre otros (Stone, 2016). En la esfera sanitaria se ha mejorado la identificación de antecedentes de pacientes, planear su atención, detectar riesgos e integrar datos genómicos y sociodemográficos que facilitan la toma de decisiones con relación a sus tratamientos (Nunes, Miranda y Dadalto, 2022; Paladino, 2021). La realidad virtual ha posibilitado mejorar servicios de salud en áreas como la formación médica, el modelado quirúrgico, la rehabilitación y la telemedicina (Megías, 2022; Shevchuk et al., 2021). Asimismo, desarrollos de IAs reparan y recrean componentes orgánicos del ser humano (García-Gutiérrez, 2017), mejoran su condición y calidad de vida, multiplicando sus potencialidades (Floridi y otros, 2018). Por último, IAs y neurotecnologías benefician la prevención, diagnóstico y terapias orientadas a mejorar el bienestar de pacientes con trastornos neurológicos (Ienca & Andorno, 2017).

En el ámbito del cuidado, asistentes robóticos acompañan cognitiva y emocionalmente a pacientes mayores monitoreando su comportamiento y su salud (Borbón, Borbón y Laverde, 2021). Asimismo, *robots* asistentes en cirugías aumentan en los médicos la visibilidad, precisión y menor invasión en las prácticas quirúrgicas, al mismo tiempo que permiten atender a un mayor número de pacientes y ayudarles a mejorar sus diagnósticos. Por último, los desarrollos de la medicina de precisión permiten pronosticar protocolos con mayor probabilidad de éxito, de acuerdo con las características del paciente y el contexto del tratamiento (Ishengoma, 2022; Paladino, 2021).

En contraste, estos sistemas de asistencia para la toma de decisiones de tratamiento basados en redes neuronales y cirugía robótica podrían convertirse en riesgo para la salud integral de los pacientes cuando se configuran con tareas similares a los cuidadores humano (D): prescribir dosis incorrectas de medicamentos debido a posibles errores de cálculo o limitación de información (Nunes, Miranda y Dadalto, 2022; Ishengoma, 2022); interpretar el comportamiento humano de manera racional y lógica debido a sus limitaciones para interactuar con características humanas como el pensamiento crítico, habilidades interpersonales y de comunicación, empatía, inteligencia emocional e imaginación (Ishengoma, 2022); limitaciones para considerar aspectos contextuales, funcionales e históricos de los seres humanos (Paladino, 2021); deshumanización del cuidado sanitario (James et al., 2023). Por otra parte, en la relación médico-paciente se origina otra amenaza para el bienestar físico y psicológico causado por la ausencia del contacto humano y la separación de la compañía e interacción interpersonal, cuando se induce al confinamiento solitario de los pacientes (Ishengoma, 2022; Van Est & Gerritsen, 2017).

En el ámbito social, otros autores indican amenazas para la salud mental cuando se configura un acoso en línea, mediante *bots*, los cuales imitan usuarios reales y rebotan respuestas automáticas a cuentas perfiladas, mediante procesamientos de lenguaje natural (Andersen, 2021). Estos sistemas ofrecen servicios y acosan con publicidad, a partir de los datos compartidos, que se convierten en capital para alimentar la financiación y ganancias de algunas empresas: información, datos de cuentas y perfiles con gustos y hábitos se utiliza con fines comerciales, económicos y políticos (O’neil, 2021). En este marco, la configuración de sistemas de IA trasiega de lavados cerebrales a la manipulación y programación cerebral (Hernández, 2020).

La visión *phronética* (A) se expresa a visibilizar un derecho a la interacción y al contacto humano significativo (Verdegay et al., 2021; Van Est & Gerritsen, 2017), que inicie con la interrelación entre personas y no entre individuos-perfiles. A nivel personal, el derecho pretendería salvaguardar el bienestar individual y las habilidades socioemocionales, que fortalecen las interacciones sociales. Este derecho también se vincularía con el concepto de “control humano significativo” relacionado con el dominio del ser humano sobre los sistemas que atentan contra la autonomía de las personas. En definitiva, se trata de no reemplazar el contacto humano por mundos y sistemas virtuales y conservar tareas de cuidado, que exigen una participación emocional, íntima y personal (Van Est & Gerritsen, 2017): “los sistemas de IA deben diseñarse y gestionarse para que sean compatibles con los ideales de dignidad humana, derechos, libertades y diversidad cultural” (Tegmark, 2017, p.371).

Discusión

Los inevitables avances de las IAs plantean desafíos a los derechos humanos, cuya historia es una permanente respuesta a las amenazas a la dignidad humana, base de los derechos y de la determinación del ser humano como persona. Este pilar fundamental, amenazado en los escenarios del mundo físico, recibe ahora mayores riesgos en los mundos virtuales. Este primer aspecto de discusión es una voz para aquellos ámbitos llamados a pensar, recrear y defender los derechos humanos, principalmente la educación y la bioética, cuyo centro común es el ser humano y su desarrollo sostenible (UNESCO, 2005; 2015; 2019).

En segundo lugar, el desarrollo sociotecnológico configura nuevas necesidades, valores y formas de organización que inducen la urgencia de repensar los derechos humanos para preservar los ciudadanos digitales, sus libertades fundamentales y su calidad de vida (Arbulo, Olveira y Wilkins, 2018). Es necesario un nuevo prisma orientado a defender aspectos fundamentales de la vida humana, así como realidades profundas hasta ahora naturalmente protegidas: mente, interioridad e identidad.

Más allá de defender derechos humanos se deberían cultivar factores que salvaguarden lo humano y ubiquen la persona como centro de cualquier desarrollo y evaluación de las IAs (Donahoe & Metzger, 2019; Morales, 2018). Es necesario considerar el valor real de la creatividad de la IA en relación con la creatividad humana como habilidad, apreciación e imaginación (García-Sedano, 2022; *Suchacka, Muster, & Wojewoda, 2021*). De ahí la necesidad del cultivo de la creatividad, el emprendimiento y la innovación que identifican *lo humano* (Aoun, 2017), así como del ejercicio de los rituales de pensar, cuestionar, y dilucidar el futuro, acciones contrapuestas a la cultura de los datos y el cálculo (Han, 2020; Martínez, 2019). En consecuencia, es necesario repensar y recrear una educación centrada en los seres humanos para afrontar esta fuerza poderosa y fortalecer la responsabilidad personal, mediante procesos de formación integral, pensamiento crítico y autorregulación (Pedró, 2019).

Un cuarto aspecto se orienta al bioderecho frente a la incertidumbre e inseguridad para ejercer dominio sobre avances neurotecnológicos y de IAs fuertes. En diferentes contextos internacionales se realizan pronunciamientos, crean comisiones de expertos, marcos regulatorios y se plantean nuevos derechos. La motivación de la Carta derechos digitales (Gobierno de España, 2021) puede configurar actitudes que garanticen una digitalización centrada en las personas: visualizar y concientizar sobre sus impactos; anticipar iniciativas para afrontar sus consecuencias; revalidar y legitimar principios y políticas. En cualquier caso, se deberían orientar políticas para un desarrollo humano sostenible, con base en el fortalecimiento del derecho básico a interactuar con y como personas, relaciones intersubjetivas, el derecho al trato como seres humanos para un desarrollo integral (Martínez, 2019).

Por último, frente a los avances tecnológicos siempre será necesario el cultivo de la sabiduría, que anteponga el cuidado de sí mismo a la exigencia de preservar derechos y fortalezca la formación integral sobre sus aspectos legales, políticos, económicos, sociales y humanos, así como el civismo, la intimidad, el pudor, el respeto, la integridad, la identidad y autonomía, para poder exigir la preservación de los principios y derechos: “A medida que la IA sigue abriéndose paso en nuestra vida cotidiana, su propensión a interferir con los derechos humanos no hace más que agravarse” (Andersen, 2021, p.1).

Conclusiones

La investigación permitió evidenciar la dinámica del flujo de las IAs: de los ámbitos sociales, laborales, educativos, hacia las diversas dimensiones del ser humano. Esta invasión al ser humano plantea beneficios y riesgos para sus derechos fundamentales y su desarrollo integral. Los sistemas que integran dispositivos, datos y algoritmos superan su dimensión técnica y arriban a erigirse en sistemas o unidades epistemológicas con valores, intencionalidades, procesos integrales hacia fines determinados y orientados a impactar las libertades, autonomía, privacidad, bienestar e identidad de las personas, sin su consentimiento. De ahí la necesidad de fortalecer la discusión para recrearlos, repensarlos, reconfigurarlos o reformularlos.

Los resultados acerca de la relación entre IAs y derechos humanos sustentan que cualquier hermenéutica para su comprensión se debe mover entre beneficios y riesgos, entre oportunidades y desafíos. Una interpretación equilibrada permite abordar de forma integral el impacto de esta fuerza poderosa, con sus desarrollos, para el ser humano y sus derechos. No obstante, la investigación denotó limitaciones en esta interpretación debido a la velocidad de estos avances, la sutileza de sus estrategias y a la lentitud de resultados investigativos de áreas como la educación, la bioética y el bioderecho. Asimismo, por la escasez de resultados investigativos empíricos sobre los impactos de las IAs en la mente y en la *psiqué*.

Más allá de los resultados, algunos hallazgos surgidos de la revisión permiten abrir interrogantes para nuevas investigaciones: Los avances continúan en nuevas realidades de invasión al ser humano como el metaverso ¿qué nuevos desafíos se abren para los derechos humanos? ¿cómo tipificar jurídicamente los impactos de las IAs en la interioridad del ser humano? ¿qué estrategias podrían implementarse para afrontar los riesgos de estos desarrollos en actores educativos? ¿qué horizontes se abren para la bioética frente a los metaversos?

Referencias bibliográficas

Aguirre, A. y Manasía, N. (2015). Derechos humanos de cuarta generación: inclusión social y democratización del conocimiento. *Télématique*, 14(1), 2-16.

Andersen, L. (2021). Human Rights in the Age of Artificial Intelligence. *Access Now. LAGJS/Essay/DS* (E074): 1-40.
<https://www.jussempir.org/Resources/Democracy%20Best%20Practices/humanrightsintheageofartif.html> 10.01.2023

Aoun, J. (2017). *Robot-Proof. Higher Education in the Age of Artificial Intelligence*. MIT Press.

Arbulo, J. I., Bruna, O. y Wilkins, A. (2018). *Hacia la cuarta generación de Derechos humanos: ciberespacio, libertad y responsabilidad digital*. XVII Jornadas de Investigación de la Facultad de Ciencias Sociales – UdelaR. A 70 años de la Declaración universal de Derechos Humanos.

Baker, T., & Smith, L. (2019). *Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges*. London: Nesta.

Bakiner, O. (2023). The promises and challenges of addressing artificial intelligence with human rights. *Big Data & Society*, 10(2), 1-13. <https://doi.org/10.1177/20539517231205476>

Barrio, A. (2021). Génesis y desarrollo de los derechos digitales. *Revista de las Cortes Generales*, 110, 197-233. <https://doi.org/10.33426/rcg/2021/110/1572>

Barrios-Tao, H., y Díaz, V. (2024). Inteligencia artificial y emociones: Psicopolítica mediante datos y algoritmos (2015-2022). *Revista de Ciencias Sociales*, XXX(1), 251-267.

- Barrios-Tao, H., Díaz, V. y Guerra, Y. (2020). Subjetividades e inteligencia artificial: desafíos para ‘lo humano’. *Veritas*, 47(4), 81-107. <https://dx.doi.org/10.4067/S0718-92732020000300081>
- Beever, J., & Taylor, L. (2022). Bioethics of public commenting: Manipulation, data risk, and public participation in E-Rulemaking. *Bioethics*, 36, 18-24. <https://doi.org/10.1111/bioe.12941>
- Beuchot, M. (2016). *Hechos e interpretaciones. Hacia una hermenéutica analógica*. México: FCE.
- Borbón, D., Borbón, L. y Laverde, J. (2020). Análisis crítico de los NeuroDerechos Humanos al libre albedrío y al acceso equitativo a tecnologías de mejora. *Ius et Scientia*, 6(2), 135-161. <https://doi.org/10.12795/IETSCIENTIA.2020.i02.10>
- Cardon, D. (2018). *Con qué sueñan los algoritmos. Nuestra vida en el tiempo de los Big Data*. Madrid: Dado Ediciones.
- Castelblanco, W. (2021). *Algoritmos que discriminan: Derechos Humanos e Inteligencia Artificial*. Trabajo de Maestría. Universidad Católica de Colombia.
- Castellanos, J. (2019). La democracia algorítmica: inteligencia artificial, democracia y participación política. *Revista General de Derecho Administrativo*, 50, 1-32.
- Castellanos, J. (2018). Transhumanismo, algoritmos y nuevas tecnologías: avanzando en la desigualdad. *IUS ET SCIENTIA*, 4(2), 120-131.
- Colmenarejo, R. (2018). Ética aplicada a la gestión de datos masivos. *Anales de la Cátedra Francisco Suárez*, 52, 113-129.
- Comisión Europea (2020). *Libro Blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza*. COM (2020) 65. https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_es.pdf
- Comisión Europea (2018). *Algorithms and Human Rights. Study on the human rights dimensions of automated data processing techniques and possible regulatory implications*. DGI (2017) 12. <https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5>
- Cotino, L. (2019). Riesgos e impactos del Big Data, la inteligencia artificial y la robótica. enfoques, modelos y principios de la respuesta del derecho. *Revista General de Derecho Administrativo*, 50, 1-37.
- Cova-Fernández, E. (2022). Derechos Humanos y Derechos Digitales en la Sociedad de la Información. *Revista DH/ED: derechos humanos y educación*, 1(6), 61-80.
- Cruz, M. (2017). De la biopolítica a la psicopolítica en el pensamiento social de Byung-Chul Han. *Athenea Digital*, 17(1), 187-203. <https://doi.org/10.5565/rev/athenea.1782>
- De Asís, R. (2020). *Inteligencia artificial y derechos humanos*. Madrid: Universidad Carlos III.
- De La Quadra-Salcedo, T. (2018). Retos, riesgos y oportunidades de la sociedad digital. En De la Quadra-Salcedo, T. y Piñar, J. (dir.), *Sociedad digital y Derecho* (pp. 21-85). Madrid: Boletín Oficial del Estado.
- Donahoe, E., & Metzger, M. (2019). Artificial intelligence and human rights. *Journal of Democracy*, 30(2), 115-126. <https://doi.org/10.1353/jod.2019.0029>
- Echeverría, J. (2016). Ciberpersonas y cibermasas en las cibersociedades. *Acento*. <https://acento.com.do/opinion/ciberpersonas-cibermasas-las-cibersociedades-8408618.html>

- Espinosa, L. (2018). La mano y el algoritmo. Una antropología compleja ante los desafíos tecnológicos del presente. *Araucaria*, 20(40), 109-136. <https://doi.org/10.12795/araucaria.2018.i40.05>
- Faralli, C. (2015). Protección de los derechos y las nuevas tecnologías. El caso de la genética en el ámbito de los derechos de la era tecnológica. *Opción*, 31(76), 15-39.
- Feldman, Z., & Sandoval, M. (2018). Metric power and the academic self: Neoliberalism, knowledge and resistance in the British university. *TripleC*, 16(1), 214-233. <https://doi.org/10.31269/triplec.v16i1.899>
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, Ch., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People –An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, 28(4), 689-707. <https://doi.org/10.1007/s11023-018-9482-5>
- Foucault, M. (2009). *El yo minimalista y otras conversaciones*. Buenos Aires: La marca editora.
- Galparsoro, J. y Pérez, R. (2018). Revolución digital y psicopolítica. Algunas consideraciones críticas a partir de Byung-Chul Han, Foucault, Deleuze y Nietzsche. *SCIO: Revista de Filosofía*, 14, 251-275. https://doi.org/10.46583/scio_2018.14.493
- García-Sedano, M. (2022). Sobre la autonomía, la creatividad y las consideraciones éticas de la inteligencia artificial en el arte contemporáneo. *H-ART. Revista de historia, teoría y crítica de arte*, 12, 17-40.
- García-Gutiérrez, J. (2017). ¿Cómo mejorar al ser humano? Un análisis de las tecnologías convergentes desde la antropología pedagógica. *Pedagogía e Vita*, 75, 94-105.
- Ghotbi, N. (2023). The Ethics of Emotional Artificial Intelligence: A Mixed Method Analysis. *Asian Bioethics Review*, 15, 417-430. <https://doi.org/10.1007/s41649-022-00237-y>
- Gobierno de España. Ministerio de Asuntos Económicos y Transformación Digital (2021). *Carta Derechos digitales*. https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/140721-Carta_Derechos_Digitales_RedEs.pdf
- González, J., Buñuel, J. y Aparicio, R. (2011). Listas guía de comprobación de revisiones sistémicas y metaanálisis: declaración PRISMA. *Evidencias en Pediatría*, 7(97), 1-6.
- Guerrero, R. (2023). Derechos humanos de cuarta generación y las tecnologías de la información y de la comunicación. En Comisión Estatal de Derechos Humanos Jalisco. *Derechos fundamentales a debate* (pp. 137-149).
- Han, B-Ch. (2014). *Psicopolítica. Neoliberalismo y nuevas técnicas de poder*. Barcelona: Herder.
- Han, B-Ch. (2020). *La desaparición de los rituales*. Barcelona: Herder.
- Hernández, J. R. (2020). *Algoritmos y su incidencia en la conducta humana. Big data, algoritmos, decisiones y seres humanos*. Universidad Autónoma de México.
- Hill, R. (2016). What an algorithm is? *Philosophy and Technology*, 29(1), 35-59. <https://doi.org/10.1007/s13347-014-0184-5>
- Huici, V. y Davila, A. (2016). Del Zoon Politikón al Zoon Elektronikón. Una reflexión sobre las condiciones de la sociedad a partir de Aristóteles. *Política y Sociedad*, 53(3), 757-772. http://dx.doi.org/10.5209/rev_POSO.2016.v53.n3.50776

Human Rights Council. (2021). *Report of the United Nations High Commissioner for Human Rights: The right to privacy in the digital age* (Report A/HRC/48/31). <https://documents.un.org/doc/undoc/gen/g21/249/24/pdf/g2124924.pdf?token=wdLnDXMUslONT1iO1N&fe=true>

Ienca, M., & Andorno, R. (2017). Towards new human rights in the age of neuroscience and neurotechnology. *Life Sciences, Society and Policy*, 13(5). <https://doi.org/10.1186/s40504-017-0050-1>

Illou, E. (2007). *Intimidaciones congeladas. Las emociones en el capitalismo*. Madrid: Katz.

Ishengoma, F. (2022). Artificial Intelligence in Digital Health: Issues and Dimensions of Ethical Concerns. *Innovación y Software*, 3(1), 81-108. <https://doi.org/10.48168/innosoft.s8.a48>

Ivanii, O., Kuchuk, A., & Orlova, O. (2020). Biotechnology as Factor for The Fourth Generation of Human Rights Formation. *Journal of History Culture and Art Research*, 9(1), 115-121. <https://doi.org/10.7596/taksad.v9i1.2540>

Ivanson, H., Altenhofen, B., Cannesson, M., & Canales, C. (2023). The Ethics of Big Data and Artificial Intelligence in Perioperative Medicine: Is Unregulated AI Already at the Bedside? *Current Anesthesiology Reports*, 13, 196-201. <https://doi.org/10.1007/s40140-023-00572-2>

James, Y., Rogers, W., Braunack, A., Frazer, H., Win, K., Houssami, N., Degeling, Ch., Semsarian, Ch., & Carter, S. (2023). Utopia versus dystopia: Professional perspectives on the impact of healthcare artificial intelligence on clinical roles and skills. *International Journal of Medical Informatics*, 169, 104903. <https://doi.org/10.1016/j.ijmedinf.2022.104903>

Jangwan, N., Ashraf, G., Ram, V., Singh, V., Alghamdi, B. S., Abuzenadah, A., & Singh, M. (2022). Brain augmentation and neuroscience technologies: Current applications, challenges, ethics and future prospects. *Frontiers in Systems Neuroscience*, 16, 1000495. <https://doi.org/10.3389/fnsys.2022.1000495>

Jobin, A., Ienca, M., & Vayena, E. (2019). Artificial Intelligence: the global landscape of ethics guidelines». *Health Ethics & Policy Lab, ETH*, 1-20. <https://doi.org/10.48550/arXiv.1906.11668>

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15-25. <https://doi.org/10.1016/j.bushor.2018.08.004>

Karabekmez, M. (2021). Data Ethics in Digital Health and Genomics. *The New Bioethics*, 27(4), 320-333. <https://doi.org/10.1080/20502877.2021.1996965>

Klavetter, E., Cruz, J., Daniels, S., Vidal, N., & Vidal, D. (2023). Historical Medical Principles Applied to the Future of Artificial Intelligence Software. *Clinic Proceedings. Digital Health*, 1(2), 115-119. <https://doi.org/10.1016/j.mcpcdig.2023.03.005>

Kubanek, J., Brown, J., Ye, P., Pauly, K., Moore, T., & Newsome, W. (2020). Remote, brain region-specific control of choice behavior with ultrasonic waves. *Science Advances*, 6(21), 1-9.

Lane, L. (2023). Preventing long-term risks to human rights in smart cities: A critical review of responsibilities for private AI developers. *Internet Policy Review*, 12(1), 1-28. <https://doi.org/10.14763/2023.1.1697>

Larrondo, M. y Grandi, N. (2021). Inteligencia Artificial, algoritmos y libertad de expresión. *Universitas, Revista de Ciencias sociales y humanas*, 34, 177-194. <https://doi.org/10.17163/uni.n34.2021.08>

- Lim, M. (2019). Governing higher education: The PURE data system and the management of the bibliometric self». *Higher Education Policy*, 34, 238-253. <https://doi.org/10.1057/s41307-018-00130-0>
- Lupton, D. (2016). The diverse domains of quantified selves: self-tracking modes and dataveillance. *Economy and Society*, 45(1), 101-122. <http://dx.doi.org/10.1080/03085147.2016.1143726>
- Martínez, J. (2019). Inteligencia y derechos humanos en la sociedad digital. *Cuadernos Electrónicos de Filosofía del Derecho*, 40, 168-189. <https://doi.org/10.7203/CEFD.40.13846>
- Martorell, F. y Andonio, A. (2019). Asistidos por el soberano electrónico. Utopías y distopías de la inteligencia artificial. *Dilemata*, 11(30), 111-127.
- Megías, J. (2022). Derechos humanos e inteligencia artificial. *Dikaiosyne*, 37, 139-163.
- Morales, P. (2018). Entre el prisma discursivo y el ciberhumanismo: algunas reflexiones sobre Derechos Humanos de cuarta generación. *Franciscanum*, 60(169), 39-86.
- Morán, A. (2021). Responsabilidad penal de la Inteligencia Artificial (IA). ¿La próxima frontera? *IUS*, 15(48), 289-323. <https://doi.org/10.35487/rius.v15i48.2021.706>
- Nguyen, Ch. (2020). Echo chambers and epistemic bubbles. *Episteme*, 17(2), 141-161. <https://doi.org/10.1017/epi.2018.32>
- Nunes, C., Miranda, R. y Dadalto, L. (2022). Desafíos bioéticos del uso de la inteligencia artificial en los hospitales. *Revista Bioética*, 30(1). <http://dx.doi.org/10.1590/1983-80422022301509ES>
- O'neil, C. (2018). *Armas de destrucción matemática. Cómo el Big Data aumenta la desigualdad y amenaza la democracia*. Madrid: Capitán Swing.
- Page, M., McKenzie, J., Bossuyt, P., Shamseer, L., Tetzlaff, J., Akl, J., Brennan, S., Chou, R., Glanville, J., Grimshaw, J., Hróbjartsson, A., ... Alonso, S. (2021). Declaración PRISMA 2020: una guía actualizada para la publicación de revisiones sistemáticas. *Revista española de cardiología*, 74(9), 790-799. <https://doi.org/10.1016/j.recesp.2021.06.016>
- Paladino, M. (2021). Cuidado e inteligencia artificial: una reflexión necesaria. *Persona y Bioética*, 25(2), 1-11. <https://doi.org/10.5294/pebi.2021.25.2.8>
- Pariser, E. (2017). *El filtro burbuja*. Barcelona: Penguin Random House.
- Pedró, F. (dir.), (2019). *Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000366994>
- Perepolkin, S., Perepolkin, D. & Averianova, M. (2021). Typology of the fourth generation of human rights. *Journal of International Legal Communication*, 1(1), 91-101. <https://doi.org/10.32612/uw.27201643.2021.1>
- Pérez, A. (2018). Las generaciones de derechos humanos ante el desafío posthumanista. En De la Quadra-Salcedo, T. y Piñar, J. (dir.). *Sociedad digital y Derecho* (pp. 137-155). Madrid: Boletín Oficial del Estado.
- Pérez, J., Deligianni, F., Ravi, D. & Guang, Y. (2017). *Artificial Intelligence and Robotics*. EPSRC UKRAS Network. <https://www.ukras.org/publications/white-papers/artificial-intelligence-robotics/>
- Piñar, J. (2018). Identidad y persona en la sociedad digital. En De la Quadra-Salcedo, T. y Piñar, J. (dir.). *Sociedad digital y Derecho* (pp. 95-111). Madrid: Boletín Oficial del Estado.

- Rallo, A. (2010). El derecho al olvido y su protección. *Telos*, 85, 104-108.
- Razmetaeva, Y., Barabash, Y., & Lukianov, D. (2022). The Concept of Human Rights in the Digital Era: Changes and Consequences for Judicial Practice. *Access to Justice in Eastern Europe*, 3(15), 41-56. <https://doi.org/10.33327/AJEE-18-5.3-a000327>
- Robles, M. (2020). La gobernanza de la inteligencia artificial: contexto y parámetros generales. *Revista electrónica de Estudios Internacionales*, 39, 1-27. <https://doi.org/10.17103/reei.39.07>
- Saura, G. y Bolívar, A. (2019). Sujeto académico neoliberal: Cuantificado, digitalizado y bibliometrificado. *REICE*, 17(4), 9-26. <https://doi.org/10.15366/reice2019.17.4.001>
- Shevchuk, O., Bululukov, O., Lysodyed, O., Mamonova, V., & Matat, Y. (2021). Human right to virtual reality in the healthcare: legal issues and enforcement problems. *Juridical Tribune*, 11, 302-315. <https://doi.org/10.24818/TBJ/2021/11/SP/03>
- Sibilia, P. (2006). *El hombre postorgánico. Cuerpo, subjetividad y tecnologías digitales*. Buenos Aires: FCE.
- Stone, P. (dir.). (2016). *Artificial Intelligence and life in 2030*. Stanford University. https://ai100.stanford.edu/sites/g/files/sbiybj9861/f/ai_100_report_0831fnl.pdf
- Suárez, J. (2014). *El derecho al olvido, base de tutela de la intimidad: Gestión de los datos personales en la red*. *Telos*, 97, 34-42.
- Suchacka, M., Muster, R., & Wojewoda, M. (2021). Human and machine creativity: social and ethical aspects of the development of artificial intelligence. *Creativity Studies*, 14(2), 430-443. <https://doi.org/10.3846/cs.2021.14316>
- Taylor, J., Harp, J., & Elliot, T. (1991). Neuropsychologists and neurolawyers. *Neuropsychology*, 5(4), 293-305. <https://doi.org/10.1037/0894-4105.5.4.293>
- Tegmark, M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. New York: Alfred Knopf.
- Toboso, M. y Aparicio, M. (2019). Entornos de funcionamientos robotizados. ¿Es posible una robótica inclusiva? *Dilemata*, 11(30), 171-185.
- Torres, J. (2018). De generaciones a escuelas: una nueva forma de concebir a los derechos humanos. *Hojas y Hablas*, 16, 97-112. <https://doi.org/10.29151/hojas-yhablas.n16a7>
- UNESCO (2019). *Consenso de Beijing sobre la inteligencia artificial y la educación*. <https://unesdoc.unesco.org/ark:/48223/pf0000368303>
- UNESCO (2015). *Replantear la educación: ¿Hacia un bien común mundial?* <https://unesdoc.unesco.org/ark:/48223/pf0000232697>
- UNESCO (2005). *Declaración universal sobre Bioética y derechos humanos*. https://unesdoc.unesco.org/ark:/48223/pf0000146180_spa?posInSet=1&queryId=eea0c6b3-c534-4cb7-aecf-f178c59c1ae1
- Valdés, E. (2015). Bioderecho, daño genético y derechos humanos de cuarta generación. *Boletín Mexicano de Derecho Comparado*, 48(144), 1197-1228.
- Van Est, R., & Gerritsen, J. (2017). *Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality*. The Hague: The Rathenau Institute.

<https://www.rathenau.nl/sites/default/files/2018-02/Human%20Rights%20in%20the%20Robot%20Age-Rathenau%20Instituut-2017.pdf>

Van Veen, Ch. (2018). Artificial Intelligence: What's Human Rights Got To Do With It?» *Data & Society: Points*. <https://medium.com/datasociety-points/artificial-intelligence-whats-human-rights-got-to-do-with-it-4622ec1566d5>

Verdegay, J., Lamata, M., Pelta, D., y Cruz, C. (2021). Inteligencia artificial y problemas de decisión: la necesidad de un contexto ético. *Suma de Negocios*, 12(27), 104-114. <https://doi.org/10.14349/sumneg/2021.V12.N27.A2>

Villalba, J. (2020). Algor-ética: la ética en la inteligencia artificial. *Blockchain inteligencia artificial*, 1, 21-40. <https://doi.org/10.24215/25916386e062>

Yuste, R., Genser, J., & Herrmann, S. (2021). It's Time for Neuro-Rights. New Human Rights for the Age of Neurotechnology. *Horizons*, 18, 154-164. <https://www.perseus-strategies.com/wp-content/uploads/2021/03/Neuro-Rights-Horizons-Winter-2021.pdf>

Zawacki, O., Marín, V., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education –Where are the educators? *International Journal of Educational Technology in Higher Education*, 16(39). <https://doi.org/10.1186/s41239-019-0171-0>