


Principios éticos para el desarrollo de la inteligencia artificial y su aplicación en los sistemas de salud

Ethical Principles for the Development of Artificial Intelligence and its Application in Health Care Systems

Jorge E. LINARES SALGADO

Universidad Nacional Autónoma de México
lisjor@unam.mx
 <https://orcid.org/0000-0003-1877-5675>

Recibido: 31/05/2022. Revisado: 06/06/2022. Aceptado: 14/06/2022

Resumen

Se plantean cuatro principios fundamentales y diez principios éticos para los sistemas de inteligencia artificial (SIA) en general y su aplicación en la salud pública. Se exponen y comentan los principios de la *Declaración de Montreal para el desarrollo responsable de la inteligencia artificial* (2018) en que se basa esta propuesta, así como de la *Recomendación sobre la ética de la inteligencia artificial* de la UNESCO (2022). La pandemia del COVID-19 ha demostrado la necesidad de construir un sistema global de salud, así como de reacción coordinada ante las próximas pandemias. Los principios éticos aplicados a los SIA pueden servir para disminuir la disparidad y las fallas de los sistemas de salud. La integración de SIA en salud de distintas regiones del mundo posibilitaría una acción global más eficiente, pero si se realiza desde el marco de los principios (bio)éticos que aquí se plantean: responsabilidad, precaución, autonomía y justicia, así como el principio de preservación de las decisiones humanas. La IA puede ayudar a desplegar progresivamente un sistema global de atención a la salud de cobertura universal y remota que atienda uno de los más importantes reclamos de justicia global: el derecho humano de atención a la salud.

Palabras clave: sistemas de inteligencia artificial; ética; bioética; salud global; principios éticos.

Abstract

Four fundamental principles and ten ethical principles are proposed for artificial intelligence systems (AIS) in general and their application in public health. The *Montreal Declaration for the Responsible Development of Artificial Intelligence* (2018) on which this proposal is based is presented and commented on, as well as the UNESCO *Recommendation on the Ethics of Artificial Intelligence* (2022). The COVID-19 pandemic has shown the need to build a global health care system, as well as a coordinated response to the coming pandemics. The ethical principles applied to AIS can serve to reduce disparity and failures of health systems. The integration of AIS in health in different regions of the world would enable a more efficient global action, but if it is carried out from the framework of the (bio)ethical principles that are raised here: responsibility, precaution, autonomy and justice, as well as the principle of preservation of human decisions. AI can help progressively to implement a global health care system with universal and remote coverage that responds to one of the most important demands for global justice: the human right to health care.

Keywords: artificial intelligence systems; ethics; bioethics; global health; ethical principles.

1. Introducción

Los sistemas de inteligencia artificial (SIA) utilizan algoritmos y procesan grandes volúmenes de información, ya sea en un software o aplicaciones, artefactos físicos interconectados o robots; son capaces de replicar la cognición, acción y aprendizaje humanos, así como planificar y tomar decisiones, por lo que constituyen una nueva modalidad de agencia colectiva.

Los sistemas de inteligencia artificial realizan tareas cada vez más complejas equivalentes o superiores a las de la inteligencia humana. Pueden procesar inmensos conjuntos de datos, calcular y predecir patrones, aprender y adaptar sus respuestas a situaciones cambiantes y novedosas, reconocer y clasificar objetos, fenómenos y rostros, interpretar, traducir y comunicar lenguajes orales y escritos, o bien realizar diagnósticos mediante el análisis de síntomas. Por ello, superan ya a la inteligencia humana en su capacidad de procesar *big data* de forma casi instantánea, por lo que tenderán a ser considerados como sistemas de *superinteligencia* (Bostrom, 2017). Los SIA pueden extenderse a distintos campos de actividad para potenciar la interacción y la intervención en la realidad social y en los ecosistemas.

Existe consenso en que la IA puede generar amplios beneficios; en particular, en servicios de atención a la salud, educación, comunicación o investigación

científica, así como mejorar las instituciones de justicia y gobernabilidad, reforzar la seguridad pública o mitigar el impacto de las actividades humanas en el medio ambiente.

Sin embargo, el desarrollo de la inteligencia artificial plantea desafíos éticos y riesgos sociales de amplio alcance, en la medida en que se extenderá su uso mundialmente. Entre sus efectos más inquietantes destacan: los SIA pueden restringir y subordinar la autonomía de las personas, afectar su capacidad de decisión y razonamiento, influir en la política y en la toma de decisiones, obstruir derechos como el de la privacidad y la intimidad, suplantar a muchos trabajadores en tareas automatizables, exacerbar las desigualdades sociales y económicas, reforzar la opacidad e incomprendibilidad de los sistemas tecnológicos o dañar aún más el clima y el medio a causa de la proliferación de artefactos inteligentes. Así, la IA puede afectar la autonomía, la responsabilidad, la justicia y el bienestar humanos.

Por otra parte, la experiencia derivada de la pandemia del COVID-19 ha mostrado la necesidad de construir un sistema global de atención a la salud, así como de reacción eficiente y coordinada ante las próximas pandemias. Para ello será indispensable contar con SIA interconectados globalmente. Los principios éticos de los SIA pueden aplicarse para reducir la desigualdad y desproporción de los sistemas de salud. Las bases de datos de salud de distintas regiones del mundo posibilitarían una acción global más eficiente y oportuna, pero regida por los principios (bio)éticos primordiales que aquí se plantean. Asimismo, la IA puede ayudar a desplegar progresivamente un sistema global de atención a la salud de cobertura universal y remota que atienda uno de los reclamos de justicia más importante, ante el déficit de profesionales sanitarios en diversos países: el derecho humano de atención a la salud en cualquier parte del mundo.

2. El problema de la disolución de la responsabilidad humana

La IA incrementará progresivamente su capacidad de *agencia intencional*, replicando y reemplazando a la agencia humana, generando el problema de la *ausencia o disolución de la responsabilidad ética en los sistemas tecnológicos*. Los SIA, robots y máquinas de IA son *agentes*, pero no *agentes morales* pues carecen (y no es claro cómo podrían obtenerlas) de conciencia, libre voluntad, emociones subjetivas y capacidad de juicio moral ponderado y contextualizado (Coeckelbergh, 2021, 111). Sin embargo, sus acciones deberán ser consideradas intencionales o deliberadas. Luciano Floridi ha planteado que, dado que los SIA son agentes racionales, los principios éticos para la IA pueden ser los mismos cuatro principios clásicos de la bioética: beneficencia, no-maleficencia, autonomía y justicia (Beauchamp y Childress, 1999), más la adición de un quinto principio de “explicabilidad” que surge en el desarrollo de la IA, el cual implica tanto la inteligibilidad y transparencia de los SIA, como la rendición de cuentas y atribución de responsabilidades por sus efectos (Floridi, 2021).

La adición del principio de “explicabilidad”, que incorpora tanto el sentido epistemológico de “inteligibilidad” (como respuesta a la pregunta “¿cómo funciona?”) como en el sentido ético de “responsabilidad” (como respuesta a la pregunta “¿quién es responsable de la forma en que funciona?”), es la pieza crucial que falta en el rompecabezas ético de la IA. Complementa los otros cuatro principios: para que la IA sea benéfica y no maléfica debemos ser capaces de comprender el bien o el daño que realmente está haciendo a la sociedad, y de qué manera; para que la IA promueva y no limite la autonomía humana, nuestra “decisión sobre quién debe decidir” debe estar informada por el conocimiento de cómo actuaría la IA en lugar de nosotros; y para que la IA sea justa, debemos saber a quién responsabilizar en caso de un resultado grave y negativo, lo que requeriría a su vez una comprensión adecuada de por qué surgió este resultado (Floridi, 2021, 12).¹

No obstante, el principal riesgo es que los SIA *no sean en el futuro tan inteligentes*, es decir, sensibles, empáticos y deliberativos como pueden ser todavía la mayoría de los seres humanos. Si a partir de la lógica del crecimiento económico y productivo se tiende a sustituir con SIA las labores que antes solo podían realizar los humanos (como la atención directa a personas) los errores y daños que ocasionen los SIA pueden ser muy graves, pero de difusa responsabilidad por la dificultad de imputarla a *alguien*. Dichas fallas tendrían que ser reparadas por las empresas y los gobiernos que autorizaran su uso, mientras que las personas afectadas tendrán que reclamar su derecho a recibir indemnizaciones y compensaciones, lo cual puede ser un calvario en algunos sistemas jurídicos.

Así, uno de los problemas éticos más serios del desarrollo de los SIA es la disolución o ausencia de imputabilidad y responsabilidad agencial por los efectos sociales y ambientales de esta novedosa tecnología que controlará a otros sistemas ya existentes (energía, comunicación, educación, salud, justicia penal). Por ello, se ha planteado que la expansión mundial de la IA exigirá mayor transparencia y “explicabilidad”, que se traduce en la rendición de cuentas de quienes (seres humanos de carne y hueso) sean los responsables de su diseño, construcción y operación, mediante regulaciones estatales más efectivas de las que han existido en el mundo industrial con anteriores productos y máquinas “no inteligentes”. Así que la tradicional tendencia a disolver la responsabilidad ética industrial, por efecto del “desfase prometeico” que planteó Günther Anders (2011), no puede continuarse en el despliegue de los SIA. Este sería un error de fatales consecuencias, pues podría conducirnos al colapso de varios de los sistemas socio-técnicos.

¹ Todas las traducciones de las citas han sido elaboradas con la ayuda de la IA del traductor de Microsoft.

² El “desfase prometeico” de Anders es la fórmula que indica la desproporción y desajuste entre nuestras capacidades tecnológicas y la falta de conciencia y responsabilidad moral por las consecuencias negativas o contraproducentes del desarrollo tecnológico-industrial, derivando en una disolución de la responsabilidad humana, individual y colectiva (Anders, 2011).

3. Valores ante los riesgos de los SIA

En suma, ¿cuáles son los principales riesgos éticos de los SIA? Estos sistemas pueden:

- Debilitar y devaluar las capacidades cognitivas y agenciales humanas en la medida en que, en contraparte, incrementen nuestras capacidades de acción colectiva.
- Restringir la autonomía de las personas, afectar su capacidad de decisión y responsabilidad en la medida en que aumente la capacidad y la autonomía de la agencia artificial.
- Influir en la política y en la toma de decisiones empleando sesgos discriminatorios en sus algoritmos, en la medida en que perdamos el control de los efectos de los SIA.
- Obstruir derechos fundamentales como el de la privacidad y la intimidad y, en esa medida, reducir la cohesión social.
- Suplantar y reemplazar a muchos trabajadores humanos en tareas automatizables y, en esa medida, menguar la capacidad de cooperación y colaboración humanas.
- Reforzar la opacidad e ininteligibilidad de los sistemas tecnológicos en la medida en que los SIA se vuelvan más complejos e intrincados.
- Exacerbar las desigualdades sociales y económicas en la medida en que los SIA sean dejados en manos del mercado capitalista mundial.
- Dañar el clima y el medio ambiente con un uso desproporcionado de energía, en la medida en que proliferen artefactos “inteligentes” que impliquen mayor contaminación y emisiones de carbono.

En contraste, el desarrollo y uso de sistemas de IA debería basarse en estos valores que protegen la primacía cognitiva y práctica de los seres humanos (Coeckelbergh, 2021,152):

- Preservación de la dignidad humana y la integridad corporal.
- Autonomía (capacidad de agencia personal).
- Justicia y equidad.
- Solidaridad, responsabilidad, rendición de cuentas.
- Decisiones democráticas sujetas a escrutinio y revisión.
- Seguridad y confiabilidad.
- Protección irrestricta de la privacidad y la intimidad.

- Transparencia e inteligibilidad de los sistemas de IA.
- Sustentabilidad ambiental.

4. Principio fundamental para los SIA: la irremplazabilidad de lo humano

La IA no deberá nunca tomar decisiones importantes (decisiones de vida o muerte) o de gran impacto y alcance social o ambiental reemplazando a los seres humanos que sean profesionalmente competentes, ni cancelar o eludir el debate y la deliberación democrática y la participación ciudadana en la toma de decisiones. La IA no debe convertirse en una nueva forma de tecnocracia dogmática y absolutista. El peor efecto ético que puede producir la IA es que los seres humanos renuncien, voluntaria o forzosamente, a su función esencial de ser agentes intencionales que deben deliberar para tomar decisiones en cada circunstancia de su vida. En todo caso, la IA debe ser un auxiliar para la toma de decisiones, cuando la inteligencia humana no tiene capacidad suficiente o cuando puede sesgarse fácilmente.

[...] está claro que la autonomía humana debe ser promovida, tanto como que la autonomía de las máquinas debe ser limitada e intrínsecamente reversible, cuando la autonomía humana tenga que ser protegida o restaurada (pensemos en el caso de un piloto capaz de desactivar el piloto automático y recuperar el control total de una aeronave). Lo anterior introduce una noción que puede definirse como *meta-autonomía*, o modelo de *decisión de delegación*. Los seres humanos deben conservar el poder de decidir qué decisiones tomar, ejerciendo la libertad de elección cuando sea necesario y cediéndola en los casos en que razones de importancia primordial, como la eficacia, puedan prevalecer sobre la pérdida de control en la toma de decisiones. Pero cualquier delegación también debe seguir siendo en principio revocable, adoptando como última garantía el poder de la *decisión de decidir de nuevo* (Floridi, 2022, 84).

5. Principios éticos generales para los SIA

Los principios éticos que aquí propongo para el desarrollo de los SIA y su aplicación en los diversos ámbitos sociales³ se basan en los cuatro principios clásicos de la bioética (beneficencia, no-maleficencia, autonomía y justicia) y en los principios de precaución y responsabilidad tecnológica, así como de remediación y protección ambiental, que han sido desarrollados a partir de planteamientos como el del *Principio de responsabilidad* de Hans Jonas (Jonas, 1995). Pueden

³ Estos son: administración y gobierno, desarrollo industrial, diseño y producción artefactual, comercio, educación, comunicación, sistemas de justicia, atención a la salud, protección del medio ambiente, investigación científica, difusión cultural, sistemas de salud, principalmente.

integrase en cuatro principios generales: defensa de la *autonomía individual y comunitaria*, preservación de la *responsabilidad humana, precaución y prevención* de riesgos y daños, *justicia distributiva y retributiva* de beneficios y riesgos (Linares, 2008).

Entre diversas propuestas que han surgido recientemente, además de la *Recomendación ética* de la UNESCO, destaca la del grupo del Centre de Recherche en Éthique de Montreal, encabezado por la Universidad de Montreal, con la colaboración de las universidades de McGill, Laval, Concordia y de Québec: *Declaración de Montreal para un desarrollo responsable de la inteligencia artificial* (CRE, 2018)⁴. Me basaré en esta Declaración para apuntalar la propuesta de implementar los principios éticos necesarios para la regulación pública de los SIA.

En particular, el caso de los sistemas de salud es de relevancia como un ejemplo concreto de un ámbito de aplicación de los SIA, pues dichos sistemas tenderán a emplear cada vez más los SIA en el diagnóstico, investigación y prevención, así como atención e intervención médicas, recolectando una masiva cantidad de información que proviene de datos privados y confidenciales sobre el estado de salud y diversos padecimientos de las personas, así como hábitos y formas de vida, por lo que aquí surge una segunda cuestión ética primordial:

6. Segundo principio fundamental para los SIA

Dado que se dificulta la anonimización de los datos electrónicos, es imperativo que toda la información procesada mediante técnicas de minería informática no sea difundida ni vendida a empresas privadas lucrativas, ni a compañías que operan los seguros de gastos médicos ni a cualquier otra compañía que lucre con la información confidencial. Los Estados deben asumir, de una vez por todas, el compromiso de proteger *efectivamente* los derechos de propiedad, integridad y privacidad de la información personal y, especialmente, la de carácter sanitario. La propiedad de dichos datos debe ser inalienable e inviolable. De lo contrario, seguiremos en la ruta de acrecentar los riesgos que implica la no regulación que favorece todo tipo de usos ilícitos y criminales de la información personal.

Ahora bien, sigue siendo válido donar datos personales mediante consentimiento informado para fines de investigación científica (no mercadotécnica); pero en el caso de fines comerciales, las personas deberían, por lo menos, recibir un pago o compensación por la información que cada una decida entregar. Así que el multimillonario negocio de las empresas globales de informática como Google o Meta (Zuboff, 2020) debe ser regulado mediante el principio de protección de la identidad de los datos personales.

⁴ Puede consultarse directamente en: <https://www.declarationmontreal-iaresponsable.com/la-declaration>

Así, los datos personales son el oro de nuestro tiempo y, entre ellos, los datos de salud, biométricos y sociodemográficos, entre otros, son considerados por la legislación como categorías especiales de datos que requieren la más alta protección porque lo dicen todo sobre nosotros y podrían utilizarse con fines no deseados y dar lugar a discriminaciones encubiertas, con profundas implicaciones para la libertad de las personas y de las generaciones futuras. La posesión de conjuntos de datos personales por parte de terceros, bien sea de iniciativa pública o privada, puede afectar a nuestros derechos en función de los usos y conferir a estos terceros un extraordinario poder sobre nosotros, situación que pasa inadvertida para la mayoría de nosotros (OByD, 2020, 39).

7. El impostergable desarrollo de un sistema global de atención a la salud

La información de salud que se procese digitalmente debe ser utilizada, prioritariamente, para investigaciones epidemiológicas y para mejorar los sistemas de atención, el desarrollo de técnicas de diagnóstico y tratamientos farmacológicos y terapéuticos *in situ* y a distancia. El objetivo fundamental para los sistemas públicos de salud debe ser fortalecer su capacidad preventiva y precautoria sobre los riesgos de salud para mejorar el uso racional de los recursos de atención (siempre escasos), tanto humanos como materiales y digitales. Lo mismo puede decirse de la aplicación de los SIA a cualquier otro ámbito público como el educativo, el cultural o de diseño y producción de bienes necesarios para la subsistencia.

Los sistemas públicos de salud podrían utilizar SIA para construir bases de datos nacionales (y también globales o coordinadas entre diversas bases nacionales de datos) a partir de expedientes e historiales clínicos digitales para implementar teleconsultas, teleradiológicos preventivos, estudios de riesgos de salud (biológicos, ambientales y sociales) y educación para la salud mediante aplicaciones digitales. Pero toda la información personal que se recabe mediante dichos SIA, públicos o privados, debe ser adecuadamente protegida para evitar que sea empleada con fines de intrusión o uso indebido de la información confidencial.

Los SIA deben ayudar a mejorar la *relación humana* entre médicos y pacientes en el diagnóstico y seguimiento de casos y nunca suplantar por completo (a menos que sea de extrema necesidad por ausencia de *personal* médico) la intervención y decisión humanas en la atención a la salud. Como hemos dicho, *la IA nunca debe tomar decisiones cruciales de vida o muerte sobre las y los pacientes*, sino ser siempre un elemento auxiliar para las decisiones médicas. En el caso en el que se lograran implementar sistemas robóticos de cirugía, revisión o diagnóstico, estos deberán ser supervisados constantemente por humanos, y sus fallas deberán ser responsabilidad de los operadores y empresas que los manufacturen. La atención remota permitirá expandir un sistema global de salud de carácter universal

y cosmopolita, pero deberá lograrse reduciendo sus inherentes riesgos. Este debe ser un objetivo ético que combatiría la injusticia y la desigualdad mundial en el acceso a los servicios sanitarios.

Los SIA ciertamente pueden conformar una medicina más *personalizada*, eficiente y de menor riesgo farmacológico, potenciar la capacidad de predicción de los efectos adversos de los medicamentos, ayudar a mejorar los hábitos saludables para incrementar el bienestar social, desarrollar sistemas de predicción y gestión de pandemias en tiempo real (como hubiera sido el caso de la pandemia del COVID-19), así como estudiar y monitorear diversos factores sociales de riesgo que inciden en las enfermedades y padecimientos más comunes.

Sin embargo, no existen aún en muchos países de desarrollo medio o bajo infraestructuras públicas para almacenar, administrar y compartir datos personales de salud. Como señala el Observatorio de Bioética y Derecho de la Universidad de Barcelona (OByD, 2020), no sería conveniente que dicha infraestructura digital se dejara en manos de las grandes compañías tecnológicas estadounidenses ya mencionadas, el anterior “imperio GAFAM” por sus siglas: Google, Apple, Facebook (ahora Meta), Amazon y Microsoft. El desarrollo de un sistema mundial de salud basado en IA debe estar a cargo de la OMS o de un consorcio internacional con recursos propios aportados por todos los Estados en la medida de sus finanzas, así como dotado de capacidad de autonomía de gestión y operación en todo el orbe.⁵

No obstante las enormes posibilidades de contribución al bienestar colectivo, los SIA en los sistemas de salud pública pueden comportar los siguientes riesgos que deben ser evitados y corregidos por el marco de principios éticos, ya que recolectarán y procesarán datos sensibles de carácter demográfico, sociocultural y socioeconómico en el análisis de síntomas y diagnósticos de padecimientos, efectividad de tratamientos y medicamentos, así como resultados curativos, reincidencias y causas de muertes. Floridi y colaboradores (Floridi *et al.*, 2021, 319-320) destacan que los SIA en la salud comportan estos riesgos principales en sus operaciones algorítmicas:

- Resultados o evidencias equivocadas, no conclusivas o inescrutables.
- Sesgos, inducción en la conducta individual (efectos “transformativos”) y resultados sociales inequitativos o discriminatorios.
- Generalización indebida y no trazabilidad de fallas y daños derivados de la IA.

⁵ ONG como *Médicos sin fronteras* representan un ejemplo de lo que podría lograr un sistema mundial de atención a la salud que usara, a la vez, SIA remotos e *in situ*, combinando personal médico y artefactos robóticos. En situaciones lamentables de guerra, como la actual invasión rusa en Ucrania, pueden ayudar a la población civil y salvar muchas vidas. Cf. Sitio web de *Médicos sin fronteras*, disponible en: <https://www.msf.org/>

8. La gestión de la información personal en los SIA de salud

Ahora bien, resulta un imperativo ético que ni los gobiernos ni las grandes empresas tecnológicas, pero tampoco las instituciones de investigación, puedan tener control pleno sobre los datos personales, y que la gestión algorítmica de estos debe someterse a criterios de transparencia para evitar la opacidad y la irregularidad que dominan actualmente en los entornos digitales, y que es favorecida por la desregulación capitalista del mercado mundial y aprovechada por redes criminales internacionales.

La mercantilización de los datos personales, en el caso de la salud pública así como en cualquier interacción social, debe ser restringida. Tampoco puede basarse solamente en la tradicional “donación altruista”, modelo que se ha supuesto en todos los proyectos de investigación biomédica, puesto que la información personal digitalizada podría ser usada con fines lucrativos sin que necesariamente produzca beneficios a los donantes (Mittelstadt y Floridi, 2016). Es decir, la era de la investigación supuestamente humanitarista, basada en la donación altruista de datos o células, debe terminar para dar paso a convenios y contratos específicos celebrados y firmados por cada persona participante en cualquier investigación. Esta prescripción es reforzada por el hecho de que ahora se dificulta en extremo anonimizar los datos personales y asegurar el consentimiento informado.

Ya que los *datos personales* son cualquier tipo de información de una persona viva que la haga identificable y, por ende, potencialmente manipulable o controlable, un aspecto crucial de seguridad digital es que deben instituirse políticas claras y eficaces que eviten la vulnerabilidad de las personas ante los riesgos de discriminación en el empleo de algoritmos que dañen los derechos o intereses legítimos de minorías, grupos sociales y personas que ya han sido marginalizadas o afectadas por los SIA. Es responsabilidad de los Estados y de las empresas evitar cualquier intento de manipulación o inducción de la conducta de las personas, aunque fuera con fines de legítimo mejoramiento de la salud pública.

En todo SIA, aplicación digital o proyecto de investigación que obtenga datos personales de salud no será suficiente con solicitar el acostumbrado “consentimiento informado”, sino que deberán implementarse mecanismos eficientes que permitan a las personas eliminarlos o “retirarlos” en cualquier momento de las bases de datos y proyectos de investigación, si consideran que han sido empleados con fines distintos a los acordados, sin que ello implique un largo proceso judicial.

Ahora bien, en situaciones de emergencia sanitaria y pandemias, será necesario establecer excepciones a la exigencia de consentimiento informado para recolectar la mayor cantidad de datos personales para vigilancia epidemiológica promovidos por las autoridades sanitarias, pero esta información deberá estar

adecuadamente protegida por los Estados. Esto es, la única excepción a la protección de datos personales y al consentimiento informado debe ser la investigación biomédica durante una pandemia nacional o mundial.

El uso de los SIA en salud pública implica dos problemas adicionales que señalan acertadamente Floridi y sus colaboradores, al respecto de las decisiones y las responsabilidades: a) el triaje para decidir quién debe ingresar a las UCI en una emergencia sanitaria (como en la pandemia del COVID-19), pues el triaje asistido por IA podría sesgarse de acuerdo con los valores introducidos en sus algoritmos por ejemplo, que priorizara solamente la edad como criterio de ingreso o rechazo), ya que el apoyo de IA en el cribado de las UCI “aumenta el número de agentes y complica las normas implicadas en estas decisiones” (Floridi *et al.*, 2021, 327); b) la dificultad de trazar y encontrar en dónde se produjeron errores en un cadena de operaciones en un sistema de salud (por ejemplo, en un hospital) en donde interactuarán agentes humanos y de IA.

El creciente uso de IA en salud está transformando la prestación de la atención médica y las implicaciones éticas de este proceso, particularmente en términos de resultados potencialmente injustos. Este proceso de transformación significa que los sistemas de salud ahora dependen de una serie dinámica, cíclica y entrelazada de interacciones entre agentes humanos, artificiales e híbridos [...]. Esto está haciendo que sea cada vez más difícil identificar los riesgos emergentes de interacción y asignar la responsabilidad, lo que plantea preocupaciones éticas con respecto a la responsabilidad moral. [...] En un sistema de salud impulsado por algoritmos, una sola herramienta de diagnóstico de IA podría involucrar a muchas personas que organizan, recopilan y procesan datos [...] lo que hace que esa asignación transparente de responsabilidad sea casi imposible. En esencia, no solo el proceso de toma de decisiones de un solo algoritmo es una caja negra, sino que toda la cadena de actores que participan en el producto final de las soluciones de IA en salud es extremadamente compleja. Esto hace que todo el ecosistema de IA en la salud sea inaccesible y opaco, lo que dificulta la responsabilidad y la rendición de cuentas (Floridi *et al.*, 2021, 328).

9. Los usos de la IA en los sistemas de salud

Los principales ámbitos de salud en los que la IA sería útil son: a) diagnóstico, análisis de patrones e investigación de causas y evolución de enfermedades; b) prevención y educación para la salud; c) atención e intervención médicas (consulta, prescripción de medicamentos, dosificación y cirugías robóticas). Para estos fines es indispensable la recolección de información masiva que proviene de datos personales y confidenciales. Con base en dicha información se pueden realizar investigaciones, seguimiento y vigilancia epidemiológicos de las principales enfermedades que afectan a la población de un país, región o a la humanidad en

su conjunto, por lo que debe prorizarse la investigación sobre las enfermedades que aquejan a los más pobres y vulnerables. Una de las posibilidades novedosas de gran relevancia es que la recolección y procesamiento de datos se ejecute en tiempo real y de formas más rápidas para obtener resultados de aplicación inmediata. Este objetivo representa por sí mismo un enorme desafío tecnológico que deben resolver los diseñadores y operadores de los SIA.

10. Tercer principio fundamental de los SIA

Los Estados deben asumir el compromiso de proteger los derechos de propiedad, integridad y privacidad de toda la información personal de carácter sanitario que recopilen y administren los SIA, y destinarla a la investigación y tratamiento de enfermedades más comunes.

11. Los principios éticos generales para los SIA

He planteado ya tres principios fundamentales para los SIA (más adelante un cuarto y último). Ahora expondré los principios éticos generales que son comunes a diferentes documentos elaborados por distintos organismos y grupos de trabajo. El de la UNESCO centra sus recomendaciones éticas en los siguientes principios (UNESCO, 2022):

- Proporcionalidad e inocuidad.
- Seguridad y protección.
- Equidad y no discriminación.
- Sostenibilidad.
- Derecho a la intimidad y protección de datos.
- Supervisión y decisión humanas.
- Transparencia y explicabilidad.
- Responsabilidad y rendición de cuentas.
- Sensibilización y educación.
- Gobernanza y colaboración adaptativas y de múltiples partes interesadas.

Dichos principios son aplicables a diversos ámbitos⁶, pero aquí me ocupo solamente de la salud pública; sin embargo, algunas de las tesis y conclusiones son extensibles a otros ámbitos públicos como la educación, la cultura, la comunicación o la investigación científica.

Una gestión prudente y responsable de la IA implicará considerarlos en conjunto como un sistema de contrapesos, de tal manera que ningún principio quede fuera del diseño y construcción de los SIA y de su utilización concreta. Las instancias de evaluación y supervisión que los Estados establezcan, así como las de ONG y sociedad civil que puedan ser auxiliares (observatorios ciudadanos), deberían basarse en estos principios, así como proponer e instaurar leyes y acuerdos internacionales para normar su desarrollo.

12. Los principios éticos de la *Declaración de Montreal de 2018* y de la *Recomendación de la UNESCO de 2022*

¿A quiénes van dirigidos los principios éticos de la IA? A los gobiernos, empresas, poderes legislativos y judiciales, agencias de regulación y evaluación, instituciones de investigación, sociedad civil organizada para realizar evaluaciones y promulgar normativas y leyes. Los principios generales para el desarrollo de los SIA que plantea la *Declaración de Montreal* son: 1. Beneficencia; 2. Autonomía; 3. Privacidad e intimidad; 4. Responsabilidad y solidaridad; 5. Participación democrática; 6. Justicia y equidad; 7. Inclusión de la diversidad; 8. Precaución; 9. Responsabilidad humana; 10. Desarrollo sustentable.

Por su parte, La *Recomendación ética* de la UNESCO propone similares principios, en los que se observa variación para ajustarse a las necesidades y funciones de los Estados: 1. Proporcionalidad e inocuidad; 2. Seguridad; 3. Equidad y no discriminación; 4. Sostenibilidad; 5. Intimidad y protección de datos; 6. Supervisión y decisión humanas; 7. Transparencia y explicabilidad; 8. Responsabilidad; 9. Educación; 10. Gobernanza y colaboración.

Como se puede apreciar, ambos conjuntos de principios coinciden en la relevancia de preservar las decisiones y responsabilidades humanas, que es la peculiaridad más destacada de la ética para la IA. Asimismo, sobresalen los principios de protección de datos personales, la sustentabilidad, la transparencia y la responsabilidad de las empresas y los Estados. Los cuatro principios generales que ya he enunciado engloban a todos ellos, pero se especificaran más adelante. Se pueden agrupar así:

⁶ Dichos ámbitos de aplicación son: 1. Gobernanza y políticas públicas, 2. Desarrollo y cooperación internacional, 3. Medio ambiente y ecosistemas, 4. Cuestiones de género, 5. Cultura, 6. Educación e investigación, 7. Comunicación e información, 8. Economía y trabajo, 9. Salud y bienestar social.

1. Responsabilidad: beneficencia y no-maleficencia, seguridad y sostenibilidad, gobernanza;
2. Precaución y prevención: explicabilidad y trazabilidad, inteligibilidad y transparencia, rendición de cuentas e imputabilidad de responsabilidades;
3. Autonomía: privacidad y confidencialidad, seguridad, preservación de la decisión humana;
4. Justicia: equidad y no-discriminación, solidaridad, inclusión de diversidad, igualdad de género y participación democrática.

Mi propuesta, basada en la *Declaración de Montreal* del Centre de Recherche en Éthique (CRE), se desplegará en diez principios generales interrelacionados (en lugar de los cinco que propone Floridi) que se exponen al final de este artículo.

¿Pero estos principios éticos pueden ser más que buenos deseos?; ¿cuál es su viabilidad en el mundo tecnológico contemporáneo? ¿Podemos asegurarnos de que, como dice la *Recomendación* de la UNESCO, “las personas deberían tener buenas razones para confiar en que los sistemas de IA pueden aportar beneficios individuales y compartidos, al tiempo que se adoptan medidas adecuadas para atenuar los riesgos”? (UNESCO, 2022, 18). Mucho me temo que no, en las condiciones actuales del sistema económico y político mundial; así que es más conveniente postular los principios éticos de forma negativa en la idea de que son preventivos y precautorios, que establecen límites que nos pueden guiar para una evaluación más precisa del desarrollo y uso de los SIA, así como para actuar oportunamente para corregir un SIA o suspenderlo si causa un daño mayor.

Los principios éticos par la IA deben funcionar como reglas de diseño y operación, así como marco normativo que delimitaría las posibles operaciones de remediación y rediseño, suspensión, desactivación o destrucción de los SIA que incumplan los principios. El problema para aplicar estas medidas correctivas es que no bastará con la acción de los gobiernos nacionales, se requerirán tratados internacionales de recíproca vigilancia y cooperación para lograr que los SIA no se conviertan en monstruos imposibles de detener y dismantelar en el caso de que sea necesario. Por ello se requerirá crear una agencia internacional e intergubernamental encargada de la supervisión y desactivación de SIA, a la manera en que funciona el Organismo Internacional de Energía Atómica creado por la Asamblea General de la ONU en 1957.⁷

De no cumplirse con estos principios éticos, las sociedades civiles, las ONG, las universidades e instituciones de investigación, los organismos internacionales

⁷ Véase el sitio web Del OIEA: Historia | OIEA (iaea.org). Cabe señalar que Corea del Norte no es estado miembro del OIEA, por lo que sus instalaciones y armas nucleares representan una grave amenaza mundial ante la opacidad y la falta de supervisión internacional.

de la ONU, tendremos la responsabilidad de dar seguimiento, señalar las fallas y violaciones a los principios, así como estudiar detalladamente los efectos sociales e individuales del uso previsiblemente global e intensivo de los SIA. Esta acción colectiva deberá presionar a las empresas y gobiernos para cumplir con los principios y actuar en consecuencia, evitando daños que pueden ser de amplio alcance. Por tanto, necesitamos establecer explícitamente un nuevo *contrato social* para el desarrollo de la IA. La mayor transparencia e inteligibilidad en el objetivo, diseño y operación son necesarias para prevenir efectos secundarios que después serán muy difíciles de desactivar o combatir.

13. Cuarto principio fundamental para los SIA

Todos los SIA deben ser diseñados y construidos de tal forma que se asegure la posibilidad efectiva de su suspensión, desactivación o destrucción por una acción humana o por otro SIA⁸. Todo sistema debe contener un mecanismo de “apagado” o “autodestrucción” en caso de que viole los principios éticos fundamentales, introyectados en sus propios algoritmos, o que dañe irreversiblemente a personas, especies vivas o ecosistemas, así sea accidentalmente.

Al respecto, la *Recomendación* de la UNESCO apunta los criterios de imputabilidad de todos los efectos de los SIA:

Deberían elaborarse mecanismos adecuados de supervisión, evaluación del impacto, auditoría y diligencia debida, incluso en lo que se refiere a la protección de los denunciantes de irregularidades, para garantizar la rendición de cuentas respecto de los sistemas de IA y de su impacto a lo largo de su ciclo de vida. Dispositivos tanto técnicos como institucionales deberían garantizar la auditabilidad y la trazabilidad (del funcionamiento) de los sistemas de IA, en particular para intentar solucionar cualquier conflicto con las normas relativas a los derechos humanos y las amenazas al bienestar del medio ambiente y los ecosistemas (UNESCO, 2022, 23).

14. Propuesta sintetizada de principios éticos para los SIA

Esta propuesta recoge el modelo de la *Declaración de Montreal* (DM) y la amplía más allá de los cuatro principios que he mencionado al inicio. Señalaré entre corchetes lo que corresponde en particular a los SIA en salud pública, con mis anotaciones adicionales. No los cito textualmente, porque intento mejorar su redacción en español y agrego o modifico algunas de las ideas expresadas en los principios fundamentales ya expuestos. El orden de presentación de los principios no indica

⁸ O, por lo menos, un mecanismo que reduzca su operación a lo menos riesgoso o a una parte no conflictiva de sus acciones y resultados.

una jerarquía axiológica, pues todos son *prima facie* e igualmente indispensables para una adecuada regulación pública de los SIA. Sin embargo, los he agrupado en orden de prioridad y alcance.

I. Principio de preservación de la responsabilidad humana (CRE, 2018, 16)

- Los seres humanos deben ser considerados, en última instancia, únicos responsables de las decisiones derivadas de las recomendaciones, indicaciones y predicciones resultantes de los SIA, así como de las acciones específicas que procedan de estos.
- En todos los ámbitos donde se debe tomar una decisión que afecta la vida, la calidad de vida o la imagen pública personal, la decisión final debe ser tomada por un ser humano y esa decisión debe ser libre de coacción e informada, inteligible y auditable o trazable.
- La decisión que implique la muerte de una persona o un animal sintiente siempre debe ser tomada por seres humanos, y la responsabilidad de esta decisión no debe transferirse a un SIA para evadir la responsabilidad. [Para los SIA en salud este es un límite infranqueable, igualmente debería serlo para los SIA aplicados en operaciones militares. Para el caso de decisiones de dar muerte a animales sintientes en vida silvestre o cautiverio, los SIA tampoco deberían actuar con independencia].
- Las personas que manipulen o alteren los SIA para cometer un crimen o un delito, o que demuestren haber actuado con negligencia u omisiones al permitir que los SIA los cometan, serán responsables, en términos de las leyes aplicables, de ese crimen o delito.⁹

Por su parte, la *Recomendación* de la UNESCO acierta en especificar para este principio que siempre debe preservarse la supervisión y la decisión humanas:

Puede ocurrir que, en algunas ocasiones, los seres humanos decidan depender de los sistemas de IA por razones de eficacia, pero la decisión de ceder el control en contextos limitados seguirá recayendo en los seres humanos, ya que estos pueden recurrir a los sistemas de IA en la adopción de decisiones y en la ejecución de tareas, pero un sistema de IA nunca podrá reemplazar la responsabilidad final de los seres humanos y su obligación de rendir cuentas. Por regla general, las decisiones de vida o muerte no deberían cederse a los sistemas de IA (UNESCO, 2022, 22).

⁹ La *Declaración de Montreal* (DM) establece que: “Cuando un SIA provoca daños o perjuicios y se demuestra que el SIA es responsable y se utilizó según lo previsto, no es razonable culpar a las personas involucradas en su desarrollo o uso” (CRE, 2018, 16). En este punto estoy en total desacuerdo, pues siempre debe ser imputable a agentes humanos la responsabilidad por los daños y perjuicios de los SIA.

II. Principio de precaución y prevención¹⁰ (CRE, 2018, 15)

Cuando el uso de un SIA ponga en peligro la salud o la seguridad públicas es necesario restringir su uso o desactivarlo precautoriamente hasta comprobar si el riesgo prevalece o se resuelve. [En los casos en que un SIA en salud implique un riesgo mayor que pueda afectar la vida de muchas personas, este debe ser suspendido o desactivado de inmediato].

- Antes de ser puestos en el mercado o ya sea que se ofrezcan de forma gratuita, los SIA deben cumplir con estrictos requisitos de confiabilidad, seguridad e integridad y ser sometidos a pruebas que no pongan en peligro la vida de las personas, dañen su calidad de vida o afecten negativamente su reputación o integridad psicológica.
- El desarrollo de los SIA debe prevenir los riesgos de uso indebido de los datos de los usuarios y proteger la integridad y confidencialidad de los datos personales mediante mecanismos eficientes de IA.
- Los errores, sesgos y defectos descubiertos en los SIA y bases de datos deben ser difundidos globalmente por instituciones públicas encargadas de la supervisión, las cuales deben señalar si estas deficiencias representan un riesgo significativo para la integridad personal o la organización social [como es el caso de la salud pública].

III. Principio de beneficencia y no-maleficencia¹¹ (CRE, 2018, 8)

- El uso de los SIA no debe afectar o disminuir la autonomía y las capacidades cognitivas de las personas, causarles daños o restringir sus libertades y derechos, a menos de que se trate de una acción específica para la prevención de un delito o detención de una persona que esté cometiendo un delito o afectación mayor a otras personas o al medio ambiente.
- Los SIA no deben afectar o ser una fuente de malestar individual, a menos que por ese medio se alcance un bienestar físico y mental superior al que no se podría alcanzar de otra manera. [En el caso de los sistemas de salud, podemos recurrir al viejo principio del “doble efecto” en forma precautoria: ningún posible beneficio en la salud pública que implique una afectación a las personas debe darse a costa de un mal mayor, individual o socialmente].

¹⁰ La DM habla de “principio de prudencia”, pero la prudencia (*phrónesis*) es un concepto ambiguo y débil para expresar un imperativo ético fundamental para la IA: la precaución y prevención a cargo de gobiernos, empresas e instituciones.

¹¹ Extiende el principio de “bienestar” de la DM a los dos principios básicos de la bioética que están coaligados: beneficencia y no-maleficencia.

- El uso de SIA no debe contribuir a aumentar el estrés y la ansiedad por causa del acoso o vigilancia permanente directa o por medio de un sistema político o policial, a menos que sea justificado por una orden judicial. Esto implica que debe evitarse un daño a la salud mental pública e individual por el uso intensivo de los SIA. [La sociedad civil debe contar con los mecanismos legales efectivos para defenderse de la intrusión invasiva de los SIA, así como medios informáticos que lo posibiliten. En el caso de la salud pública, no se justifica el uso de información de salud para persecuciones judiciales].

IV. Principio de respeto a la autonomía individual y la diversidad cultural (CRE, 2018, 9)

- Los SIA deben permitir y favorecer que las personas puedan cumplir sus propios objetivos vitales y su concepción de una vida buena, de acuerdo con los valores de su comunidad cultural y con base en el respeto a los derechos y autonomía de las demás personas. [Los SIA en la salud pública pueden promover hábitos y cuidados saludables, verificados y revisables científicamente, pero no imponer un estilo de vida único a toda la población, mucho menos derivado del modelo occidental de vida ecológicamente no sustentable].
- Los SIA no deben desarrollarse ni utilizarse para imponer un estilo de vida particular a los individuos, ya sea directa o indirectamente, mediante la implementación de mecanismos de vigilancia, inducción o de incentivos directos que puedan ser injustos o discriminatorios. [Los SIA en salud no deben ser usados para captar sujetos de investigación que no reciban beneficios directos o para entregar sus datos para fines lucrativos mediante la vigilancia de su actividad personal, ni para discriminar o dar prioridad injustificada de tratamiento a un grupo social sobre otros].
- Las instituciones públicas no deben utilizar los SIA para promover o desacreditar una concepción particular o condición específica de vida o concepción moral. [Debe prevenirse que el uso de SIA en la salud pública no estigmatice o discrimine a personas con discapacidad, padecimientos poco comunes o ya estigmatizados socialmente, como las personas con VIH, o con preferencias o identidades sexo-genéricas no heteronormativas].
- Los SIA no deben desarrollarse para difundir información sesgada o falsa, o propaganda de carácter ideológico y político, y deben diseñarse con el fin de impedir su difusión, por lo que deben estar dotados de mecanismos de detección de esta información no confiable que pueda contaminarlos: rumores, confusiones, noticias e información falsas y

sesgos mal intencionados. [Toda la información en salud que usen los SIA debe ser validada y corroborada, siempre verificable y objetable por cualquier usuario que pueda aportar evidencias y denunciar fallas y abusos].

- El desarrollo de los SIA debe evitar la creación de dependencias a través de técnicas de captación de atención o imitación de características humanas atractivas (apariencia, voz, lenguaje, etc.) de maneras que podrían causar confusión sobre si se trata de un agente de SIA o un humano. [En el uso de los SIA con fines médicos no deben utilizarse técnicas de manipulación de las preferencias o valoraciones de las personas, o incluso métodos de persuasión ideológicos y propagandísticos, aunque fuera con fines de mejoramiento de la salud pública].

V. Principio de protección a la privacidad y la intimidad (CRE, 2018, 10)

- La intimidad de los pensamientos y las emociones debe protegerse estrictamente en los usos de SIA para evitar causar daño o vulnerabilidad, especialmente en aquellas operaciones que impliquen juicios morales sobre las elecciones personales y estilos de vida. [El uso de SIA en salud pública no debe tratar de imponer concepciones morales de un grupo en particular, ni difundir valores morales que no sean plenamente justificables, universales o que violenten los derechos de otras personas].
- Las personas siempre deben tener derecho a la desconexión digital en su vida diaria; los SIA deben ofrecer explícitamente la opción de desconectarse a intervalos regulares, y no deben alentar mediante incentivos subrepticios a mantenerse conectadas. [Los SIA en salud no deben crear dependencia de uso ni incentivos para mantenerse conectado adictivamente a los SIA entregando y procesando datos de manera permanente].
- Los SIA y bases de datos médicas deben garantizar la confidencialidad de los datos y el anonimato del perfil personal, en la medida de lo posible. [Los datos de salud pertenecen legalmente a las personas y no a las empresas o instituciones públicas].
- El acceso a los SIA y a los servicios digitales por parte de las personas no debe estar condicionado a que abandonen el control o la propiedad de sus datos personales. [En los SIA de salud, las personas deben poder retirar o borrar sus datos sin tener evidencia de su mal uso o comercialización].
- Las personas deben ser libres de donar sus datos a organizaciones de investigación para contribuir al avance del conocimiento, mediante las

reglas y prescripciones del consentimiento informado. [La investigación en salud basada en SIA debe ser transparente y proteger la información confidencial de las personas].

VI. Principio de responsabilidad social y solidaridad¹² (CRE, 2018, 11)

- Los SIA no deben debilitar o impedir relaciones humanas moral y psicológicamente satisfactorias, y deben desarrollarse con el objetivo de fomentar estas relaciones y reducir la vulnerabilidad y el aislamiento de las personas, creando medios de vinculación y cooperación social, así como de conversación y debate constante en plataformas digitales.
- Los SIA deben desarrollarse con el objetivo de auxiliar a las personas en tareas complejas y debe fomentar el trabajo colaborativo entre humanos y SIA. [Los SIA en salud no deben reemplazar a la capacidad de trabajo colectivo humano, a menos que sea en una situación de emergencia o imposibilidad de la presencia de personal médico humano, como en el caso de una emergencia sanitaria].
- Los SIA no deben implementarse para reemplazar a las personas en tareas que requieren relaciones humanas de calidad y proximidad, sino que deben desarrollarse para facilitar estas relaciones o establecerlas de manera remota. [Por ejemplo, en la relación personal médico-pacientes, especialmente en temas de salud mental. Si es necesario, los SIA intervendrán en forma remota o virtual cuando no sea posible establecer relaciones personales entre los médicos y los pacientes].
- Los conocimientos científicos y técnicos, así como los tradicionales, validados científicamente, que usen o recopilen los SIA, especialmente en el ámbito de salud, no deben ser privatizados vía patentes ni apropiados indebidamente por ningún grupo social, empresa o gobierno, sino difundidos universalmente.

VII. Principio de justicia y equidad¹³ (CRE, 2018, 13)

- Los SIA deben diseñarse y operarse para no crear, reforzar o reproducir la discriminación social basada en diferencias de género, étnicas, nacionales, culturales, lingüísticas o religiosas.

¹² La DM se queda corta en este principio al reducirlo a la “solidaridad”, debe expandirse a la idea de corresponsabilidad social.

¹³ La DM solo especifica para este principio la equidad. Es necesario fortalecer la idea de justicia distributiva y retributiva.

- El desarrollo de los SIA debe ayudar a eliminar las relaciones de dominación entre grupos y personas basadas en diferencias de poder, riqueza o conocimiento, así como prevenir y evitar la violencia y la discriminación negativa. [Particular atención debe ponerse en reducir y evitar la desigualdad y la discriminación por género. La *Recomendación* de la UNESCO aborda específicamente este punto. Los SIA no deben contribuir a la desigualdad y brechas de género¹⁴, sino a visibilizarlas y disminuirlas].
- El desarrollo de los SIA debe producir beneficios sociales, políticos y económicos mediante la reducción de las desigualdades y vulnerabilidades de los grupos más desfavorecidos. [La atención mundial a la salud mediante SIA remotos podría facilitar la reducción de brecha de desigualdad mundial entre países y sectores sociales].
- El desarrollo de los SIA debe estar destinado a cumplir acciones de justicia retributiva y a generar opciones de desarrollo a los grupos sociales vulnerables o que tradicionalmente han estado marginados del desarrollo y del ejercicio de libertades y derechos. [La investigación en salud por medio de los SIA debe priorizar la atención de enfermedades más comunes y las que afectan a las personas más pobres y vulnerables].

VIII. Principio de transparencia y explicabilidad¹⁵ (CRE, 2018, 12) (UNESCO, 2022, 22-23) (Floridi, 2021, 12, 31)

- Los SIA que tomen decisiones que afecten la vida, la calidad de vida o la reputación personal deben ser inteligibles y transparentes para sus operadores y usuarios.
- Las decisiones tomadas y las acciones ejecutadas por los SIA que afecten la vida, la calidad de vida o la reputación personal siempre deben ser expresadas en un lenguaje que sea comprensible por las personas que los usan o que están sujetas a las consecuencias de su uso. [En salud,

¹⁴ La *Recomendación* de la UNESCO señala: “Entre estas brechas cabe citar la disparidad salarial entre hombres y mujeres; su representación desigual en ciertas profesiones y actividades; la falta de representación en los puestos directivos superiores, las juntas directivas o los equipos de investigación en el campo de la IA; la brecha educativa; las desigualdades en el acceso, la adopción, la utilización y la asequibilidad de la tecnología digital y de la IA; y la distribución desigual del trabajo no remunerado y de las responsabilidades de cuidado en nuestras sociedades” (UNESCO, 2022, 32).

¹⁵ Subsumo la “participación democrática” de la DM en la transparencia y la explicabilidad, dado que su definición es difusa y ambigua. ¿Qué significa participación democrática en la IA? Para unos, basta con que las personas puedan opinar; para otras, tendrían que poseer el poder de determinar el curso de los SIA, para otros evaluar y elegir entre diferentes opciones de SIA, y para otros significa el poder de evitar o desactivar SIA que no cumplan con los principios éticos.

los SIA no deben usar lenguaje técnico críptico, sino siempre inteligible para cualquier persona con educación básica. Asimismo, los SIA deben ser plurilingüísticos para abarcar las principales lenguas usuales de cada nación o región del planeta].

- El descubrimiento de errores operativos en los SIA, efectos inesperados o indeseables, violaciones de seguridad y fugas de datos debe informarse perentoriamente a las autoridades pertinentes, a las partes interesadas y las personas afectadas. [Será necesario para los SIA en salud la creación de organismos e instituciones que supervisen su funcionamiento mediante auditorías informáticas exhaustivas, y que posean la autoridad para suspender o desactivar un SIA, si ello no tiene consecuencias inmediatas de mayor daño social].
- Para los SIA públicos que tienen un impacto significativo en la vida de las personas, estas deben tener la oportunidad para deliberar sobre los parámetros sociales y algoritmos de los SIA, sus objetivos y los límites de su uso, así como elegir entre diferentes opciones. [Los SIA en salud deben ser desarrollados y evaluados con participación ciudadana plural. Una de las formas elementales de la democracia es que las personas tengan al alcance diversas opciones para elegir las que mejor les convengan, por lo que debe evitarse la monopolización de SIA privados o públicos].
- Cualquier persona que utilice un servicio de IA debe saber si una decisión que le concierne o que le afecta fue tomada por un SIA o por un robot, y no por un ser humano. Las decisiones deben poder ser explicadas, tanto en el sentido epistémico (cómo se generaron en el SIA) como en sentido moral (quiénes son responsables de ellas). [En el caso de los SIA en salud, debe evitarse siempre que la IA tome decisiones por sí misma. Aunque la trazabilidad pueda ser intrincada, las auditorías a los SIA deben poder encontrar el origen de las decisiones].
- Cualquier usuario de un servicio que emplee chatbots debe poder identificar fácilmente si está interactuando con un SIA o con una persona humana. [Solo en casos de necesidad extrema, los SIA en salud pueden operar con chatbots o agentes robóticos que tomen decisiones al respecto de la atención y tratamiento de pacientes].

IX. Principio de inclusión de la diversidad e igualdad de género (CRE, 2018, 14)

- El desarrollo y uso de los SIA no debe conducir a la homogeneización de la sociedad a través de la estandarización de comportamiento y opiniones típicas del mundo occidental (sobre todo, lo que no es sustentable). [Los SIA en salud no deben promover un estilo único o

hegemónico de vida, y tienen que combatir mitos y sesgos en las ideas sociales sobre la salud. Debe valorarse y rescatarse el conjunto de conocimientos tradicionales terapéuticos que sean validados por métodos e investigaciones científicas¹⁶].

- Desde el diseño de los algoritmos, el desarrollo y despliegue de los SIA debe tener en cuenta la diversidad de expresiones culturales presentes en la sociedad, así como la igualdad de género y la diversidad sexo-génerica.
- Los SIA deben evitar el uso de los datos individuales para encasillar a las personas en un perfil de usuario, estereotipar su identidad personal o confinarla a un “filtro burbuja¹⁷”, que restrinja y limite sus posibilidades de desarrollo y actividad personal, especialmente en campos como la educación, la cultura, la justicia [o la salud pública].
- Los SIA no deben desarrollarse ni utilizarse con el objetivo de limitar la libre expresión de ideas o la oportunidad de valorar opiniones diversas, condiciones esenciales de una sociedad plural democrática. [Para el ámbito de la salud, no deben impedir la expresión y difusión de conocimientos tradicionales de carácter y eficacia terapéutica validada ni obstaculizar críticas y cuestionamientos a los sistemas de salud, al dispositivo epistémico-médico hegemónico y a la industria farmacéutica. Los SIA deben promover la indagación sobre el desarrollo farmacológico, aplicando el mismo principio de explicabilidad a la industria].

X. Principio de desarrollo sustentable (CRE, 2018, 17)

- El hardware y la infraestructura digital, así como los centros de datos, deben funcionar con la mayor eficiencia energética posible y mitigar las emisiones de carbono durante todo su ciclo de vida. Ello implica promover, mediante los SIA, investigación científica para mejorar su propio desempeño hasta alcanzar el objetivo final de emisiones cero.

¹⁶ Por ejemplo, los SIA en salud deberían promover las dietas vegetarianas y veganas, por razones de salud y ecológicas, así como combatir los mitos sociales que consideran que comer carne es indispensable para una buena salud o para obtener proteínas necesarias para el cuerpo. La información científica actual es que las dietas no cárnicas, bien equilibradas, no son deficientes; y en cambio, favorecen al medio ambiente, reducen riesgos de enfermedades cardiovasculares y de distintos tipos de cáncer.

¹⁷ La expresión “filtro burbuja” hace referencia a la información “filtrada” que le llega a una persona en internet o en una aplicación informática, como las redes sociales o los motores de búsqueda, que ofrecen resultados personalizados, segmentados y reducidos para los usuarios. Esto puede aislar a las personas (en “burbujas”) porque ya no pueden acceder a la información común y a los catálogos completos de datos.

- Los SIA no deberían convertirse, en lo inmediato, en una nueva fuente de crecientes emisiones de carbono a la atmósfera. [Los Estados deben tomar las medidas necesarias para evitarlo o disminuir su efecto negativo, priorizando los SIA más esenciales, como los dedicados a la salud pública].
- El hardware de los SIA, su infraestructura digital y los artefactos de IA deben tener como objetivo generar la menor cantidad de residuos materiales, así como proporcionar procedimientos de mantenimiento, reparación y reciclaje de acuerdo con los principios de la economía circular y la reducción de la huella de carbono en los niveles máximos posibles.
- El hardware de los SIA, su infraestructura digital y los artefactos de IA deben minimizar su impacto en los ecosistemas y en la biodiversidad, especialmente con respecto a la extracción de recursos y la eliminación final del equipo cuando alcanza el final de su vida útil.

15. Conclusión

La reformulación negativa (precautoria y preventiva) de los diez principios éticos generales para la IA, es decir, en términos de límites que deben ser vigilados por toda la sociedad, nos permite tener un marco regulatorio más realista y eficiente para evaluar y dar seguimiento puntual al desarrollo de los SIA. En lo particular, en el ámbito de la salud pública el uso de SIA en este marco ético puede tener consecuencias favorables para mejorar los diagnósticos y medidas de seguimiento y tratamiento de los padecimientos más comunes, epidémicos y pandémicos. Asimismo, el uso delimitado en un marco ético global de los SIA en la salud pública puede crear la infraestructura digital para hacer frente a las próximas pandemias, así como esto es lo más importante construir un sistema global de atención a la salud que llegue a todas las personas en todas partes del mundo, especialmente a las regiones más apartadas y a las poblaciones más precarizadas.

Un sistema global de atención básica de la salud recopilará y procesará toda la información necesaria disponible, de forma rápida y segura, y podrá llevar remotamente métodos probados científicamente a todas las poblaciones, contribuyendo a la distribución más racional de medicamentos y tratamientos en donde más se requieren. Pero, al mismo tiempo, puede recuperar técnicas terapéuticas tradicionales de diversas culturas y grupos originarios para difundirlas universalmente, impidiendo su privatización vía patentes, impulsando una atención a la salud multicultural y plural en beneficio de toda la humanidad.

Referencias bibliográficas

- Anders, Günther (2011). *La obsolescencia del hombre, vol. 1.*. Valencia: Pre-Textos.
- Beauchamp, Tom L., James Childress (1999). *Principios de ética biomédica*. Barcelona: Masson.
- Bostrom, Nick (2017). *Superintelligence. Paths, Dangers, Strategies*. Oxford: Oxford University Press.
- Centre de Recherche en Éthique (CRE), Université de Montréal (2018). *Déclaration de Montréal pour un développement responsable de l'intelligence artificielle*. https://www.declarationmontreal-iaresponsable.com/_files/ugd/ebc3a3_28b2dfe7ee13479caaf820477de1b8bc.pdf?index=true
- Coeckelbergh, Mark (2021). *AI Ethics*. Kindle ed. Cambridge, MA.: The MIT Press.
- Floridi, Luciano (2022). *Etica dell'intelligenza artificiale. Sviluppì, opportunità, sfide*. Milán: Raffaello Cortina.
- Floridi, Luciano (ed.) (2021). *Ethics, Governance, and Policies in Artificial Intelligence*. Switzerland: Springer.
- Floridi Luciano y Josh Cowls (2021a). A Unified Framework of Five Principles for AI in Society. En Luciano Floridi (Ed.), *Ethics, Governance, and Policies in Artificial Intelligence* (pp. 5-18). Switzerland: Springer.
- Floridi, Luciano *et al.* (2021b). The Ethics of AI in Health Care: A Mapping Review. En Luciano Floridi (Ed.), *Ethics, Governance, and Policies in Artificial Intelligence* (pp. 313-346). Switzerland: Springer.
- Jonas, Hans (1995). *El principio de responsabilidad. Ensayo de una ética para la civilización tecnológica*. Barcelona: Herder.
- Linares, Jorge E. (2008). *Ética y mundo tecnológico*. México: FCE.
- Mittelstadt, Brent D., Luciano Floridi (eds.) (2016). *The Ethics of Biomedical Big Data*. Switzerland: Springer.
- Observatorio de Bioética y Derecho (OByD), Universidad de Barcelona (2020). *Pautas para evaluar proyectos de investigación de innovación en salud que utilicen tecnologías emergentes y datos personales*. http://www.bioeticayderecho.ub.edu/sites/default/files/documents/doc_evaluar-proyectos-salud.pdf
- UNESCO (2022). *Recomendación sobre la ética de la inteligencia artificial*. https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa
- Zuboff, Shoshana (2020). *La era del capitalismo de la vigilancia*. Barcelona: Paidós.