

# Inteligencia artificial para el avance de los sistemas de salud. Posibles aportes y retos<sup>1</sup>

## Artificial intelligence for the advancement of health systems. Possible contributions and challenges

CLARA BERMÚDEZ-TAMAYO *Escuela Andaluza de Salud Pública  
Observatorio de Salud y Medio Ambiente de Andalucía (OSMAN). Escuela Andaluza  
de Salud Pública, Granada, España  
CIBER de Epidemiología y Salud Pública (CIBERESP), España*

JAIME JIMÉNEZ-PERNET *Escuela Andaluza de Salud Pública*

### Resumen

La creciente disponibilidad de grandes conjuntos de datos y de nuevos métodos analíticos, sumado a los avances simultáneos en la infraestructura de las tecnologías de la información (TIC) y la cobertura poblacional de la tecnología móvil, han motivado la esperanza de que la IA permita abordar los retos sanitarios futuros. La IA puede utilizar algoritmos sofisticados para “aprender” características de un gran volumen de datos sanitarios y utilizar los conocimientos obtenidos para ayudar a la práctica clínica. También puede estar dotada de capacidades de aprendizaje y autocorrección para mejorar su precisión en función de la información recibida. Los dispositivos de IA utilizados en el ámbito de la salud se pueden dividir en dos categorías principales: las técnicas de aprendizaje automático (Machine Learning - ML) que analizan datos estructurados como los datos de imagen, genéticos y electrofisiológicos, que intentan agrupar los rasgos de los pacientes o inferir la probabilidad de los resultados de la enfermedad, y los métodos de procesamiento del lenguaje natural (PNL) que extraen información de datos no estructurados como notas clínicas/revistas médicas para complementar y enriquecer los datos médicos estructurados. Las limitaciones actuales de la iHealth y la eHealth en general incluyen los costes financieros, las barreras culturales, lingüísticas y de alfabetización, los problemas de suministro de energía, la seguridad de los datos y los problemas de privacidad. Los puestos de trabajo en el sector sanitario con más probabilidades de ser automatizados parecen ser aquellos en los que parte de sus tareas implican el manejo de información digital, como la radiología y la patología, al contrario de aquellos basados en las relaciones humanas y el contacto directo con pacientes.

### Abstract

The increasing availability of large data sets and new analytical methods, coupled with simultaneous advances in information technology (ICT) infrastructure and mobile population coverage, have given rise to the hope that AI will address future health challenges. AI can use sophisticated algorithms to “learn” characteristics of a large volume of health data and use the knowledge gained to support clinical practice. It may also be equipped with learning and self-correcting capabilities to improve accuracy based on the information received. AI devices used in health can be divided into two main categories: Machine Learning (ML) techniques that analyze structured data such as image, genetic, and electrophysiological data, which attempt to group patient traits or infer the probability of disease outcomes, and natural language processing (NLP) methods that extract information from unstructured data such as clinical notes/medical journals to supplement and enrich structured medical data. Current limitations of iHealth and eHealth in general include financial costs, cultural, language and literacy barriers, power supply problems, data security, and privacy issues. Healthcare jobs most likely to be automated appear to be those where part of their tasks involves the management of digital information, such as radiology and pathology, contrary to those based on human relationships and direct contact with patients.

### Palabras Clave

sanidad; inteligencia artificial; aprendizaje automático; procesamiento del lenguaje natural; TIC; e-Salud

### Keywords

healthcare; artificial intelligence; machine learning; natural language processing; ICT; e-Health

<sup>1</sup> Estudio realizado en el seno del Proyecto de Investigación B-SEJ-213-UGR18 “Retos de la modernización de la asistencia sanitaria en Andalucía: Cohesión interterritorial, envejecimiento y revolución digital. Proyectos I+D+i del Programa Operativo FEDER 2018. Junta de Andalucía-Consejería de Economía y Conocimiento. Universidad de Granada

## 1. INTRODUCCIÓN

En los últimos años, la Inteligencia Artificial (IA) es citada de manera recurrente como una de las tecnologías más prometedoras para hacer avanzar los sistemas de salud y mejorar la provisión de servicios sanitarios (1). El *World Economic Forum* plantea que el desarrollo intensivo de la IA supondría la “cuarta revolución industrial” con implicaciones transformadoras y globales, y al igual que las revoluciones que la precedieron, esta tendría el potencial de aumentar los niveles de ingresos globales y mejorar la calidad de vida de las poblaciones, y por tanto, su bienestar (2).

La creciente disponibilidad de grandes conjuntos de datos y de nuevos métodos analíticos, sumado a los avances simultáneos en la infraestructura de las tecnologías de la información (TIC) y la cobertura poblacional de la tecnología móvil, han motivado la esperanza de que la IA permita abordar los retos sanitarios futuros.

Nos referimos a la escasez de personal sanitario, el aumento del gasto sanitario, y la debilidad de los sistemas de vigilancia de la salud pública. Los sistemas sanitarios podrían ahorrarse los costes de transporte de profesionales en muchos casos, y la logística y las cadenas de suministro globales serían más eficiente y el coste del comercio disminuiría, lo que podría abrir nuevos mercados.

La IA puede utilizar algoritmos sofisticados para “aprender” características de un gran volumen de datos sanitarios y, a continuación, utilizar los conocimientos obtenidos para ayudar a la práctica clínica. También puede estar dotada de capacidades de aprendizaje y autocorrección para mejorar su precisión en función de la información recibida. Un sistema de IA puede ayudar a los profesionales sanitarios proporcionándoles información clínica actualizada procedente de revistas, libros de texto y prácticas clínicas para informar sobre el cuidado adecuado de los pacientes (3). Además, teóricamente un sistema de IA podría ayudar a reducir los errores diagnósticos y terapéuticos que son inevitables en la práctica clínica (4). Por otra parte, un sistema de IA puede extraer información útil de un gran número de pacientes para hacer inferencias en tiempo real para alertar de riesgos sanitarios y estimar resultados de salud (5).

Las intervenciones basadas en la IA han demostrado, por ejemplo, que pueden informar la toma de decisiones clínicas para reducir la carga de trabajo de los profesionales sanitarios y evitar desplazamientos a zonas aisladas, mejorando la eficiencia del sistema (6). También el uso de IA ha ayudado a identificar los brotes de enfermedades con mayor antelación que los enfoques tradicionales, apoyando así la planificación de programas y una formulación de políticas oportuna (7).

Muchos países, conscientes del potencial de la IA, han desarrollado o están en proceso de desplegar estrategias y políticas para promover la investigación, el desarrollo y la adopción de métodos y tecnologías de IA. Desde el sector privado también hay una intensa inversión en I+D+i en este ámbito (8,9).

En este contexto, están claras las posibilidades de la IA y su utilidad prometedora, no obstante, el uso de la IA en el ámbito de la salud plantea importantes desafíos éticos, jurídicos, comerciales y sociales de carácter transnacional (10). El uso de programas informáticos en la atención sanitaria ya suponía grandes retos para desarrolladores, gobiernos y proveedores de servicios, la IA plantea retos éticos adicionales que van más allá del ámbito de los reguladores tradicionales y los agentes que intervienen en los sistemas de atención sanitaria (1,11). El uso apropiado para mejorar la salud, debe ser abordado adecuadamente de tal manera que se pueda preservar la autonomía de las personas y garantice un acceso equitativo a estas tecnologías. (10)

El optimismo en los beneficios potenciales de la IA o “tecno-optimismo” podría empeorar la calidad de vida de las sociedades, por ejemplo, agravando la distribución desigual del acceso a la atención sanitaria por motivos geográficos, de género, de edad o de disponibilidad de dispositivos, si los países no adoptan las medidas (12). El uso inadecuado de la IA también podría perpetuar, exacerbar o profundizar los prejuicios y las disparidades en la atención sanitaria por cuestiones raciales o étnicas

con inferencias sesgadas, por análisis de datos y herramientas mal diseñadas mediante el uso de datos limitados, de baja calidad y no representativos (10).

## 2. LA INTELIGENCIA ARTIFICIAL Y SUS TECNOLOGÍAS

La inteligencia artificial trata de las técnicas que permiten a las máquinas reproducir funciones que solemos atribuir a los animales y a los humanos, como percibir, moverse, actuar sobre el mundo o planificar (13).

No hay una definición consensuada sobre IA, por el contrario, su campo se va redefiniendo constantemente a medida que ciertos temas se clasifican como ajenos a la IA y surgen otros nuevos. Además, hay que señalar que la popularidad de la IA se debe, en parte, a que se ha empezado a utilizar este término para referirse a cosas que solían llamarse de otro modo.

El Consejo de Inteligencia Artificial de la OCDE define la IA como “un sistema basado en una máquina que puede hacer predicciones, recomendaciones o decisiones que influyen en entornos reales o virtuales, para un conjunto dado de objetivos definidos por el ser humano”.(14).

La IA puede entenderse también como “un campo de estudio que combina la informática, la ingeniería y otras disciplinas relacionadas para construir máquinas capaces de comportamientos que se diría que requieren inteligencia si se observaran en los seres humanos”(15). Algunos de estos comportamientos son la capacidad de percibir imágenes visualmente, reconocer los sonidos y diálogos, traducir el lenguaje natural y aprender de la nueva información y adaptarse a ella (3).

La clasificación entre lo que es y lo que no es IA no es una dicotomía clara sí-no: mientras que algunos desarrollos son claramente IA, también hay métodos en los que interviene “algo” de IA. Así pues, en ocasiones sería más apropiado hablar de cuánto de IA tiene algo, en lugar de determinar simplemente si ese algo es o no es IA. Para diferenciar lo que es IA y lo que no, hablamos de autonomía y adaptabilidad. Autonomía, entendida como la capacidad para ejecutar tareas en situaciones complejas sin la dirección constante del usuario, mientras adaptabilidad es la capacidad para mejorar la ejecución de las tareas aprendiendo de la experiencia.

La base de la IA son los algoritmos o códigos informáticos que contienen instrucciones para un rápido análisis y transformación de datos en conclusiones, información u otros resultados. Las enormes cantidades de datos y la capacidad de analizarlos rápidamente impulsan la IA. Los distintos tipos de tecnologías de IA incluyen aplicaciones de aprendizaje automático (machine learning) como reconocimiento de patrones, procesamiento del lenguaje natural, procesamiento de señales y sistemas expertos.

### 2.1. Aprendizaje automático (Machine Learning - ML)

El aprendizaje automático permite a los algoritmos hacer predicciones y resolver problemas basándose en grandes cantidades de datos, sin ser programados explícitamente. Se refiere a sistemas que van mejorando la manera en que ejecutan una tarea dada, a medida que acumulan experiencia o datos. El aprendizaje automático se basa en el uso de técnicas de modelización estadística y matemática para definir y analizar datos. Esos patrones aprendidos se aplican después para realizar o guiar determinadas tareas y hacer predicciones

El ML se basa en el uso de algoritmos de aprendizaje, capaces de aprender y mejorar automáticamente a partir de lo que experimentan. El uso de cada técnica debe adaptarse al problema a tratar para encontrar la solución más adecuada. El ML puede clasificarse en 3 tipos de aprendizaje: supervisado, no supervisado y de refuerzo. En el aprendizaje supervisado, los algoritmos utilizan un conjunto de datos etiquetados para predecir el resultado conocido, definido por una variable “clase”. Por el contrario, el aprendizaje no supervisado trata de identificar patrones ocultos presentes en los datos sin retroalimentación. Por último, el aprendizaje de refuerzo puede considerarse un híbrido de

los dos anteriores, con el objetivo de maximizar la precisión de los algoritmos mediante el método de ensayo y error (16).

## 2.2. Aprendizaje profundo

Es un subcampo del aprendizaje automático. El aprendizaje profundo es un subconjunto del aprendizaje automático, y va más allá al utilizar múltiples capas de redes neuronales artificiales para resolver problemas complejos a partir de datos no estructurados, de forma muy similar al cerebro humano (17). La «profundidad» se refiere a la complejidad de un modelo matemático, y que el aumento de la capacidad de computación de los ordenadores modernos ha permitido a los investigadores incrementar esta complejidad hasta alcanzar niveles no solo cuantitativa, sino también cualitativamente diferentes de los anteriores.

## 2.3. Robótica

Por robótica se entiende la construcción y programación de robots que puedan operar en situaciones complejas del mundo real. En cierto sentido, es el reto definitivo de la IA, ya que precisa una combinación de múltiples ámbitos tecnológicos. Por ejemplo, visión mediante ordenador y reconocimiento del habla para detectar el entorno, procesamiento del lenguaje natural, recuperación de información y razonamiento en condiciones de incertidumbre para procesar instrucciones y predecir las consecuencias de las posibles acciones, modelado cognitivo y computación afectiva (sistemas que responden a expresiones de sentimientos humanos o que imitan sentimientos) para interactuar y trabajar con seres humanos (3).

## 3. APLICACIONES DE LA INTELIGENCIA ARTIFICIAL EN ÁMBITO SANITARIO

La Inteligencia Artificial (IA) ha abierto un nuevo camino con el potencial de abordar problemas clínicos complejos, debido a su capacidad de identificar relaciones no lineales, conexiones profundas y errores no lineales, así como la capacidad de minimizar los errores. También ha abierto la posibilidad de nuevos modelos predictivos que puedan generar estratificaciones de riesgo en un plazo más corto (por ejemplo, 2 años o menos) con una capacidad predictiva más fiable.

Es probable que los enfoques clínicos actuales hagan que una parte de la población no reciba un diagnóstico o tratamiento adecuado, reciba uno menos efectivo, o reciba un tratamiento preventivo innecesario. Esto se debe entre otras causas, a la simplificación excesiva de las complejas relaciones entre las variables implicadas y a la asunción de correlaciones lineales entre los factores. Las técnicas de IA que se han aplicado en el ámbito sanitario prometen mejorar la calidad de la atención al paciente, mejorar la eficiencia, y reducir las tasas de mortalidad y morbilidad, a través de análisis y predicciones más precisas.(8). El proceso clínico se podría ver apoyado por esta nueva familia de tecnologías, a disposición de los profesionales sanitarios.

No obstante, antes de que los sistemas de IA puedan desplegarse en las aplicaciones sanitarias, deben ser “entrenados” a través de los datos que se generan en las actividades clínicas, como el cribado, el diagnóstico, la asignación de tratamientos, etc. Así, los sistemas IA aprenden sobre grupos similares de sujetos, las asociaciones entre las características de los sujetos y los resultados de interés.

Estos datos clínicos de base suelen existir, entre otros, en forma de datos demográficos, notas médicas, registros electrónicos de dispositivos médicos, exámenes físicos y laboratorio clínico e imágenes.(18) Los datos de la exploración física, notas de la anamnesis, y los resultados del laboratorio clínico los distinguimos de los datos de imagen, genéticos y electrofisiológicos porque contienen grandes porciones de textos narrativos no estructurados, que no son analizables directamente. Para ello, algunas aplicaciones de IA se centran en convertir primero el texto no estructurado en un registro médico electrónico (EMR) comprensible para la máquina. (19).

Para los próximos años, en las economías más avanzadas, se espera que las tecnologías de IA desempeñen un papel importante no sólo en la prestación de asistencia, sino también en la integración de los sistemas de asistencia sanitaria y social. El registro y la combinación de múltiples tipos de datos puede aumentar la precisión del diagnóstico, mejorar la personalización de la asistencia, tratar las enfermedades mentales y físicas comórbidas, hacer un seguimiento de la experiencia del paciente en los diferentes servicios de asistencia y comprender las conexiones entre la salud y la vida cotidiana de forma más amplia (20).

Con respecto al diagnóstico y seguimiento, la evaluación clínica en línea y la observación continua a distancia ofrecen cada vez más posibilidades, mediante el uso de historias clínicas electrónicas y una ecología de flujos de datos que se entrecruzan y que se recogen a través de dispositivos de *ehealth* (21). La noción de “*iHealth*” (salud inteligente) se basa en la eHealth y la amplía mediante el uso de la autovigilancia en tiempo real dentro del entorno del paciente, junto con el procesamiento de datos y la extracción de datos para apoyar la toma de decisiones personalizada.

Por ejemplo, en el ámbito de la psiquiatría, algunos estudios han demostrado que la medición de los marcadores del estado clínico a través de la analítica de datos puede permitir a los pacientes y a los médicos detectar las señales de alerta temprana y realizar un seguimiento y una predicción precisos de las recaídas en las personas que padecen trastornos mentales graves.(22) Del mismo modo, los trabajos recientes en informática afectiva y lingüística computacional sugieren que la ansiedad, la depresión, el trastorno bipolar y la intención suicida pueden observarse y medirse a través de los datos de las formas lingüísticas ordinarias que se producen de forma “natural”, como las publicaciones y los comentarios en las plataformas de las redes sociales.(23).

Sin embargo, aunque aumentan y en algunos casos superan las capacidades de los profesionales sanitarios, la precisión diagnóstica de tales técnicas de análisis de datos, como los algoritmos de aprendizaje automático no se puede dar por sentada. En la actualidad, algunas aplicaciones IA siguen produciendo tasas de error clínicamente inaceptables (24).

En relación al tratamiento/terapia, las tecnologías de uso intensivo de datos tienen un claro potencial terapéutico. La abundancia de datos digitales ha facilitado la farmacoepidemiología y, en particular, la investigación observacional sobre la eficacia de la medicación en el mundo real.(25) Hay indicios de que las terapias basadas en entornos de realidad virtual inmersiva codiseñados y los humanos virtuales (por ejemplo, chatbots) pueden superar las barreras de comunicación y ampliar el acceso a una atención de mayor calidad (26). Sin embargo, se han notificado efectos adversos y consecuencias no deseadas (27) que plantean la necesidad de implicar a las comunidades culturalmente diversas que hablan varios idiomas, a las personas con discapacidades y problemas de comunicación, a las redes de apoyo de los cuidadores formales y a los cuidadores familiares, los cuales pueden desempeñar un papel activo en la formulación de directrices y normas para evaluar la seguridad y los resultados de los pacientes.(28)

Algunos sistemas de información pueden parecer intuitivos para los pacientes con comorbilidades, ya que no suelen separar explícitamente las necesidades de salud mental y física, ni ven sus experiencias de salud y enfermedad como una parte separada de sus vidas. Una gran proporción de los participantes en la investigación reclutados en un entorno de atención de urgencias (más del 70%) han aceptado que se vincule su actividad en las redes sociales con sus historias clínicas electrónicas, aunque entre los usuarios más jóvenes de redes sociales se observa una mayor aprobación .(29)

El modo en que se recogen, cotejan e interpretan los datos constituye un reto común para la integración de las tecnologías digitales en la atención habilitada digitalmente. Destacan las complejidades asociadas a la propiedad de los datos, el valor económico de los datos personales y la monetización de los datos o las prácticas de “intercambio de conocimientos” entre diversas empresas e instituciones(30). La participación pública en estas decisiones podría desempeñar un

papel fundamental a la hora de ganarse la confianza del público en la intersección de los intereses privados y públicos en competencia. También podría informar sobre las estrategias para garantizar la interoperabilidad de las tecnologías digitales de salud con los registros sanitarios electrónicos y otras formas de datos digitales personales.

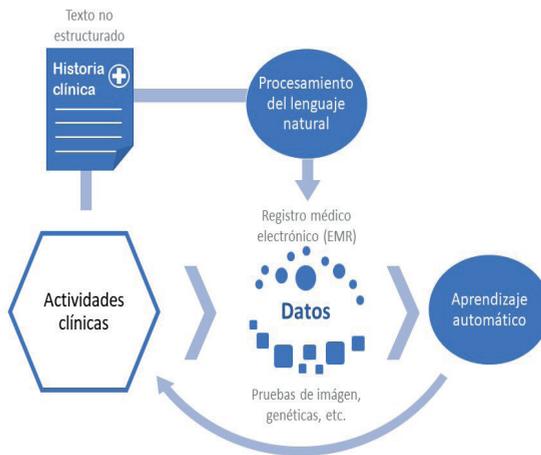
**3.1. Dispositivos de IA utilizados en el ámbito de la salud**

Los dispositivos de IA en salud, se pueden dividir en dos categorías principales. La primera categoría incluye las técnicas de aprendizaje automático (ML) que analizan datos estructurados como los datos de imagen, genéticos y electrofisiológicos. En las aplicaciones clínicas, los procedimientos de ML intentan agrupar los rasgos de los pacientes o inferir la probabilidad de los resultados de la enfermedad.(18)

La segunda categoría incluye los métodos de procesamiento del lenguaje natural (PNL) que extraen información de datos no estructurados como notas clínicas/revistas médicas para complementar y enriquecer los datos médicos estructurados. Los procedimientos de PLN tienen como objetivo convertir los textos en datos estructurados legibles por la máquina, que luego pueden ser analizados por técnicas de ML.(31)

La figura 1 describe el proceso desde la generación de datos clínicos, que indica el proceso de enriquecimiento de datos mediante PLN y el análisis de datos mediante ML, hasta la toma de decisiones clínicas. El proceso empieza y termina con las actividades clínicas; por muy potentes que sean las técnicas de IA, tienen que estar motivadas por los problemas clínicos y aplicarse para ayudar a la práctica clínica al final.

**Figura 1. Proceso de generación de datos y toma de decisiones clínicas enriquecidas con herramientas IA.**



**Elaboración propia basada en (3)**

El ML puede utilizarse tanto en salud pública como en ámbito clínico, para reclutamiento de pacientes, extracción de conocimientos de los registros narrativos y clasificación con respecto a diferentes objetivos como el diagnóstico o los riesgos. Estas aplicaciones se pueden clasificar en las categorías descritas en la tabla 1. El apoyo al diagnóstico de enfermedades transmisibles y no transmisibles ha surgido como uno de los principales usos de la IA.

**Tabla 1. Aplicaciones de IA en el ámbito clínico y la salud pública**

	<b>Tipo de inteligencia artificial</b>	<b>Ejemplos</b>
Diagnóstico	<ul style="list-style-type: none"> <li>• Sistema experto</li> <li>• Machine learning</li> <li>• Procesamiento de lenguaje natural</li> <li>• Procesamiento de señales (signal processing)</li> </ul>	Se aplicaron métodos de aprendizaje automático y procesamiento de señales a las radiografías digitales de tórax para identificar casos de tuberculosis y de tuberculosis resistente a los medicamentos
Evaluación del riesgo de mortalidad y morbilidad	<ul style="list-style-type: none"> <li>• Data mining</li> <li>• Machine learning</li> <li>• Procesamiento de señales</li> </ul>	Para cuantificar el riesgo de la gravedad del dengue, se aplicaron algoritmos de aprendizaje automático a conjuntos de datos administrativos de un hospital de atención terciaria en Tailandia
Predicción de brotes de enfermedades y vigilancia	<ul style="list-style-type: none"> <li>• Data mining</li> <li>• Machine learning</li> <li>• Procesamiento de lenguaje natural</li> <li>• Procesamiento de señales (signal processing)</li> </ul>	Se utilizaron datos de teledetección y algoritmos de aprendizaje automático para caracterizar y predecir los patrones de transmisión del virus del Zika a nivel mundial
Planeación y políticas de salud	Expert planning; machine learning	Se aplicaron modelos de aprendizaje automático a datos administrativos de Sudáfrica para predecir la duración de la estancia de los trabajadores sanitarios en comunidades desatendidas

*Basado en Schwalbe N, Wahl B. Artificial intelligence and the future of global health. Lancet. 2020 May 16;395(10236):1579-1586.*

### 3.2. Aprendizaje automático (Machine Learning - ML) en ámbito de la salud

Como se ha comentado, el ML construye algoritmos de análisis de datos para extraer características de los datos. Los datos de entrada de los algoritmos de ML incluyen los “rasgos” del paciente y los resultados clínicos de interés. Los rasgos de un paciente suelen incluir datos de referencia, como la edad, el sexo, los antecedentes de enfermedad, así como datos específicos de la enfermedad, como imágenes de diagnóstico, expresiones genéticas, pruebas, resultados de exámenes físicos, síntomas clínicos, medicación, etc. Además de los rasgos, en la investigación clínica se suelen recoger los resultados médicos de los pacientes. Estos incluyen indicadores de la enfermedad, tiempos de supervivencia de los pacientes y niveles cuantitativos de la enfermedad, por ejemplo, el tamaño de los tumores.

Dependiendo de si se incorporan los resultados, los algoritmos de ML pueden dividirse en dos grandes categorías: aprendizaje no supervisado y aprendizaje supervisado. El aprendizaje no supervisado es bien conocido para la extracción de características, mientras que el aprendizaje supervisado es adecuado para el modelado predictivo mediante la construcción de algunas relaciones entre los rasgos del paciente (como entrada) y el resultado de interés (como salida). Más recientemente, se ha propuesto el aprendizaje semi-supervisado como un híbrido entre el aprendizaje no supervisado y el supervisado, que es adecuado para escenarios en los que falta el resultado para ciertos sujetos.

El clustering y el análisis de componentes principales (PCA) son dos de los principales métodos de aprendizaje no supervisado. El clustering agrupa a los sujetos con rasgos similares en clusters, sin utilizar la información de los resultados. Los algoritmos de clustering producen las etiquetas de clúster para los pacientes mediante la maximización y minimización de la similitud de los pacientes dentro y entre los clúster. El PCA sirve principalmente para reducir las dimensiones, especialmente cuando el rasgo se registra en un gran número de dimensiones, como el número de genes en un estudio de asociación de todo el genoma. El PCA proyecta los datos en unas pocas direcciones de componentes principales, sin perder demasiada información sobre los sujetos. A veces, se puede utilizar primero el PCA para reducir la dimensión de los datos, y luego utilizar el *clustering* para agrupar los sujetos.

Por otro lado, el aprendizaje supervisado considera los resultados de los sujetos junto con sus rasgos, y pasa por un cierto proceso de entrenamiento para determinar las mejores salidas asociadas a las entradas que más se acercan a los resultados en promedio. Normalmente, las fórmulas de salida varían en función de los resultados de interés. Por ejemplo, el resultado puede ser la probabilidad de padecer un determinado evento clínico, el valor esperado de un nivel de enfermedad o el tiempo de supervivencia esperado.

### 3.3. Aprendizaje profundo

El aprendizaje profundo es una extensión moderna de la técnica clásica de redes neuronales. Se puede considerar el aprendizaje profundo como una red neuronal con muchas capas. El rápido desarrollo de la informática moderna permite al aprendizaje profundo construir redes neuronales con un gran número de capas, lo que es inviable para las redes neuronales clásicas. Así, el aprendizaje profundo puede explorar patrones no lineales más complejos en los datos. Otra razón de la reciente popularidad del aprendizaje profundo se debe al aumento del volumen y la complejidad de los datos. Una clara mayoría del aprendizaje profundo se utiliza en el análisis de pruebas de imagen, lo que tiene sentido dado que las imágenes son naturalmente complejas y de gran volumen (16).

Un ejemplo de estas aplicaciones está en el uso de aprendizaje profundo para el cribado mamario. Se pretende solventar algunas de las limitaciones actuales, como reducir la carga de trabajo al personal de radiología o eliminar la fatiga o el diagnóstico subjetivo. Sin embargo, la evidencia actual es aún insuficiente para juzgar la precisión de IA en los programas de cribado del cáncer de mama (32). Los sistemas de IA alcanzan menos especificidad que la doble lectura de radiólogos, en los programas de cribado, aunque las aplicaciones IA basadas en aprendizaje profundo pueden alcanzar una alta sensibilidad, por lo que podría ser adecuado para clasificar a las mujeres que deben recibir una revisión radiológica.

### 3.4. Procesamiento del lenguaje natural

Dar sentido al lenguaje humano ha sido un objetivo de investigación de la IA desde la década de 1950. Este campo, la PNL, incluye aplicaciones como el reconocimiento del habla, el análisis de textos, la traducción y otros objetivos relacionados con el lenguaje. Hay dos enfoques básicos: la PNL estadística y la semántica. La PNL estadística se basa en el aprendizaje automático (redes neuronales de aprendizaje profundo) y ha contribuido a un reciente aumento de la precisión del reconocimiento. Requiere un gran "corpus" o cuerpo de lenguaje del que aprender.

En el ámbito de la sanidad, las aplicaciones dominantes de la PNL consisten en la creación, comprensión y clasificación de la documentación clínica y la investigación publicada. Los sistemas de PLN pueden analizar notas clínicas no estructuradas sobre pacientes, preparar informes (por ejemplo, sobre exámenes radiológicos), transcribir interacciones con pacientes y realizar IA conversacional.

## 4. DESAFÍOS Y LIMITACIONES DE LA IA

Como hemos visto antes, un sistema de IA básico suele integrar un componente ML para manejar datos estructurados (imágenes, datos EP, datos genéticos) y un componente NLP para la minería de textos no estructurados. Los algoritmos más sofisticados deben ser entrenados a través de los datos sanitarios para desarrollar en el sistema la capacidad de ayudar a los profesionales sanitarios con el diagnóstico de la enfermedad o las sugerencias de tratamiento.

Aprovechando la gran cantidad de datos con abundante información, se espera que la IA ayude a estudiar cuestiones clínicas mucho más complicadas, pero mucho más cercanas a la vida real, lo que conduce a una mejor toma de decisiones en el tratamiento. Recientemente, personal investigador en diferentes ámbitos ha empezado a trabajar en esta dirección y han obtenido resultados iniciales prometedores (8,33).

Las limitaciones actuales de la iHealth y la eHealth en general incluyen los costes financieros, las barreras culturales, lingüísticas y de alfabetización, los problemas de suministro de energía (por ejemplo, en comunidades remotas), la seguridad de los datos y los problemas de privacidad.(30)

#### 4.1. Aspectos generales

Aunque las tecnologías de IA están atrayendo una gran atención en la investigación médica, su aplicación en la vida real sigue encontrando obstáculos. El primer obstáculo proviene de la normativa. La normativa actual carece de normas para evaluar la seguridad y la eficacia de los sistemas de IA. Para superar esta dificultad, la Administración de Alimentos y Medicamentos de los Estados Unidos (FDA) realizó un primer intento de proporcionar una guía para evaluar los sistemas de IA (34). Esta clasifica los sistemas de IA como “productos de bienestar general”, que están regulados de forma poco estricta siempre que los dispositivos se destinen únicamente al bienestar general y presenten un bajo riesgo para los usuarios. Una segunda orientación, justifica el uso de pruebas en el mundo real para evaluar el rendimiento de los sistemas de IA. Por último, la guía aclara las normas para el diseño adaptativo en los ensayos clínicos, que se utilizarían ampliamente para evaluar las características de funcionamiento de los sistemas de IA.

El segundo obstáculo es el intercambio de datos. Para que funcionen bien, los sistemas de IA necesitan ser entrenados (continuamente) por los datos de los estudios clínicos. Sin embargo, una vez que un sistema de IA se despliega tras el entrenamiento inicial con datos históricos, la continuación del suministro de datos se convierte en una cuestión crucial para el desarrollo y la mejora del sistema. El entorno sanitario actual no ofrece incentivos para compartir los datos del sistema. Sin embargo, se está llevando a cabo una revolución sanitaria para estimular el intercambio de datos, que ha sido favorecido por la pandemia por COVID19 (11).

#### 4.2. Implicaciones para el personal sanitario

Se ha prestado mucha atención a la preocupación de que la IA provoque la automatización de los puestos de trabajo y el desplazamiento sustancial de la mano de obra. Algunos trabajos han estimado que el 35% de los puestos de trabajo del Reino Unido podrían desaparecer debido a la IA en los próximos 10 a 20 años (35). Otros estudios han sugerido que, si bien es posible una cierta automatización de los puestos de trabajo, una serie de factores externos, además de la tecnología, podrían limitar la pérdida de puestos de trabajo, como el coste de las tecnologías de automatización, el crecimiento y el coste del mercado laboral, los beneficios de la automatización más allá de la simple sustitución de la mano de obra, y la aceptación normativa y social.(36)

La limitada incursión de la IA en el sector hasta ahora, y la dificultad de integrar la IA en los flujos de trabajo clínicos, han sido en cierto modo responsables de la falta de impacto en el empleo hasta la actualidad. Algunos sistemas basados en IA podrían reducir funciones o actividades repetitivas. Los puestos de trabajo en el sector sanitario con más probabilidades de ser automatizados parecen ser aquellos en los que parte de sus tareas implican el manejo de información digital, como la radiología y la patología. Todo lo contrario de aquellos puestos basados en las relaciones humanas, que tienen contacto directo con pacientes.

No obstante, los procesos clínicos para emplear el trabajo de imagen basado en la IA están muy lejos de estar listos para su uso diario. Los distintos proveedores de tecnología de imagen y los algoritmos de aprendizaje profundo tienen focos diferentes: la probabilidad de una lesión, la probabilidad de cáncer, la característica de un nódulo o su localización. Estos focos distintos harían muy difícil la incorporación de los sistemas de aprendizaje profundo a la práctica clínica actual.

Por otra parte, los algoritmos de aprendizaje profundo para el reconocimiento de imágenes requieren “datos etiquetados”: millones de imágenes de pacientes que han recibido un diagnóstico definitivo de cáncer, un hueso roto u otra patología. Sin embargo, actualmente no está claro el número

total de estos repositorios, ni los conjuntos de datos que contienen o si las imágenes radiológicas están etiquetadas o no.

También existe la posibilidad de que se creen nuevos puestos de trabajo para trabajar con las tecnologías de IA y desarrollarlas. Pero el empleo humano estático o en aumento también significa, por supuesto, que no es probable que las tecnologías de IA reduzcan sustancialmente los costes del diagnóstico y el tratamiento médico en ese plazo.

### 4.3. Implicaciones éticas

Hay una serie de implicaciones éticas en torno al uso de la IA en la asistencia sanitaria. Las decisiones sanitarias han sido tomadas casi exclusivamente por humanos en el pasado, y el uso de máquinas inteligentes para tomarlas o ayudarlas plantea cuestiones de responsabilidad, transparencia, permiso y privacidad.

En primer lugar, la capacidad de los sistemas IA debe superar la restricción de la dimensionalidad requiere conjuntos de datos voluminosos (37). Los conjuntos de datos pequeños y/o sesgados pueden dar lugar a un exceso de ajuste, lo que limita los problemas que pueden abordar los métodos actuales de aprendizaje automático. El uso de bases de datos de imágenes dermatológicas es un buen ejemplo de los retos a enfrentar. La aplicación de algoritmos para uso generalizado a nivel poblacional requiere una mayor transparencia en los repositorios y conjuntos de datos disponibles. Una revisión sistemática de los bancos de imágenes dermatológicas disponibles confirma un predominio de imágenes de lesiones cutáneas de pacientes asiáticos y pacientes blancos de EE.UU (38).

En segundo lugar, los modelos de aprendizaje automático se describen a menudo como “cajas negras” cuya opacidad impide la interpretación o la comprobación de los supuestos clave por parte de personas no expertas. Aunque los trabajos recientes suelen abordar parcialmente esta limitación, el problema persiste (39). Quizás la cuestión más difícil de abordar dadas las tecnologías actuales es la transparencia. Muchos algoritmos de IA -en particular los de aprendizaje profundo utilizados para el análisis de imágenes- son prácticamente imposibles de interpretar o explicar. Si se informa a un paciente de que una imagen ha dado lugar a un diagnóstico de cáncer, probablemente querrá saber por qué. Los algoritmos de aprendizaje profundo, e incluso el personal médico que generalmente están familiarizados con su funcionamiento, pueden ser incapaces de dar una explicación.

En tercer lugar, en algunos casos, los observadores asumen que los modelos que aprenden automáticamente de los datos son más objetivos y, por tanto, más precisos que los modelos construidos por humanos. Aunque los modelos basados en datos suelen predecir los resultados mejor que los modelos basados en la teoría, la construcción de modelos basados en datos también implica decisiones subjetivas y asunciones y parámetros, como la elección de conjuntos de datos de pruebas y evaluación, la elección de criterios de preprocesamiento y la elección de algoritmos de aprendizaje y parámetros iniciales. Estas decisiones dan lugar a sesgos y prejuicios que pueden pasar desapercibidos para los usuarios ocasionales (40). La autonomía y capacidad de elección se puede ver condicionada por sistemas que recomiendan sistemáticamente opciones basadas en decisiones previas, limitando la diversidad de opciones futura.

En cuarto lugar, la cantidad de datos suele ir en detrimento de la calidad. Esto es un problema para cualquier análisis de datos, pero puede ser especialmente pernicioso en el contexto de los métodos de aprendizaje automático que utilizan un conjunto de pruebas para estimar la precisión de la predicción en el mundo más amplio. Si los artefactos de recopilación de datos hacen que los conjuntos de prueba y reales sean demasiado similares entre sí, pero diferentes de los conjuntos de datos a los que normalmente se aplicaría el modelo, el sobreajuste puede conducir a una precisión de predicción imprevista en el mundo real (24). Por último, dado que los estudios de IA suelen requerir la vinculación de datos de uso secundario procedentes de fuentes heterogéneas, las discrepancias entre estas fuentes de datos pueden inducir sesgos, incluido el sesgo por patrones demográficos (41).

Un sistema IA puede reproducir o generar discriminación cuando los algoritmos o datos a partir de los que aprende contienen errores o sesgos que lo llevan a tratar o reconocer de forma diferenciada a grupos diversos, sin una clara justificación.

Diversos autores, como Brynjolfsson y McAfee, alertan que la IA podría generar una mayor desigualdad debido a procesos de discriminación, limitación de la autonomía o falta de transparencia, pero sobre todo por su potencial para afectar los mercados laborales. A medida que la automatización sustituya a la mano de obra en toda la economía. Hasta la fecha, los que más se han beneficiado de ella han sido los ciudadanos capaces de acceder al mundo digital.

La comunidad sanitaria mundial, incluidos varios grandes organismos financiadores, ha reconocido cada vez más la urgencia de abordar estas cuestiones para garantizar que todas las poblaciones se beneficien de los avances en materia de salud digital e IA. En mayo de 2018, la Asamblea Mundial de la Salud adoptó una resolución sobre las tecnologías digitales para la cobertura sanitaria universal y en 2019, el Grupo de Alto Nivel sobre Cooperación Digital, el Secretario General de las Naciones Unidas recomendó que *“para 2030, todos los adultos deberían tener un acceso asequible a las redes digitales, así como a los servicios financieros y de salud habilitados por medios digitales, como medio para contribuir sustancialmente a la consecución de los ODS”*. The Lancet y Financial Times creó una comisión sobre *“el futuro de la salud en 2030: crecer en un mundo digital”* y sostiene que las transformaciones digitales deben considerarse un factor determinante de la salud. La Comisión también aboga para que se reconsidere radicalmente la cuestión de las tecnologías digitales, destacando que sin un enfoque preventivo, orientado a la misión y basado en el valor de su gobernanza, las transformaciones digitales no conseguirán mejorar la salud para todos(43).

## 5. EL FUTURO DE LA IA

La IA tiene un importante papel que desempeñar en la oferta sanitaria del futuro. En forma de aprendizaje automático, es la principal capacidad que subyace al desarrollo de la medicina de precisión, que se considera un avance muy necesario en la atención sanitaria. Dados los rápidos avances de la IA en el análisis de imágenes, parece probable que la mayoría de las imágenes de radiología y patología sean examinadas en algún momento por una máquina. El reconocimiento de voz y de texto ya se emplea para tareas como la comunicación con el paciente y la captura de notas clínicas, y su uso irá en aumento.

El mayor reto de la IA en estos ámbitos sanitarios no es si las tecnologías serán lo suficientemente capaces como para ser útiles, sino garantizar su adopción en la práctica clínica diaria. Para que se produzca una adopción generalizada, los sistemas de IA deben ser aprobados por los organismos reguladores, integrarse en los sistemas de evaluación de tecnologías sanitarias, estandarizarse lo suficiente como para que productos similares funcionen de forma parecida, enseñarse a los clínicos, y actualizarse con el tiempo. Estos retos se acabarán superando, pero tardarán mucho más en hacerlo que lo que tardarán las propias tecnologías en madurar. En consecuencia, algunos se atreven a aventurar que habrá un uso limitado de la IA en la práctica clínica dentro de 5 años y un uso más amplio dentro de 10.(36)

También parece cada vez más claro que los sistemas de IA no sustituirán a los profesionales a gran escala, sino que aumentarán sus esfuerzos para atender a los pacientes. Con el tiempo, los clínicos humanos pueden orientarse hacia tareas y diseños de trabajo que se basan en habilidades exclusivamente humanas como la empatía, la persuasión y la integración de la visión global.

No cabe duda de que los sistemas de IA cometerán errores en el diagnóstico y el tratamiento de los pacientes y puede ser difícil establecer la responsabilidad de los mismos. También es probable que nos encontremos con muchos cambios éticos, médicos, laborales y tecnológicos con la IA en la asistencia sanitaria. Es importante que las instituciones sanitarias, así como los organismos gubernamentales y reguladores, establezcan estructuras para supervisar las cuestiones clave, reaccionar de forma

responsable y establecer mecanismos de gobernanza para limitar las implicaciones negativas. Se trata de una de las tecnologías más potentes y con mayores consecuencias que afectarán a las sociedades humanas, por lo que requerirá una atención continua y una política reflexiva durante muchos años.

## REFERENCIAS

1. MURPHY K, DI RUGGIERO E, UPSHUR R, WILLISON D, MALHOTRA R, CAI J. Artificial intelligence for good health: a scoping review of the ethics literature. *BMC Med Ethics*. 2021 Dec;22(1).
2. SCHWAB K. The Fourth Industrial Revolution: what it means and how to respond. World Economic Forum. 2016.
3. JIANG F, JIANG Y, ZHI H, DONG Y, LI H, MA S, *et al*. Artificial intelligence in healthcare: Past, present and future. *Stroke Vasc Neurol*. 2017 Dec;2(4):230-43.
4. MINUÉ S, BERMÚDEZ-TAMAYO C, FERNÁNDEZ A, MARTÍN-MARTÍN JJ, BENÍTEZ V, MELGUIZO M, *et al*. Identification of factors associated with diagnostic error in primary care. *BMC Fam Pract*. 2014;15(1).
5. Morley J, Machado CCV, Burr C, Cowls J, Joshi I, Taddeo M, *et al*. The ethics of AI in health care: A mapping review. *Soc Sci Med*. 2020 Sep;260.
6. Guo J, Li B. The application of medical artificial intelligence technology in rural areas of developing countries. *Heal Equity*. 2018;2:174-81.
7. Lake I, Colón-González F, Barker G, Morbey R, Smith G, Elliot A. Machine learning to refine decision making within a syndromic surveillance service. *BMC Public Health*. 2019;19:559.
8. Hakim Z, Ierasts T, Hakim I, D'Penha A, Smith KPD, Caesar MCW. Connecting Data to Insight: A Pan-Canadian Study on AI in Healthcare. *Healthc Q*. 2020 Apr;23(1):13-9.
9. CIFAR. Pan-Canadian AI Strategy. Invest in Canada. 2017.
10. WHO. Ethics and governance of artificial intelligence for health ethics and governance of artificial intelligence for health. World Health Organization. 2021.
11. El-Haddadeh R, Fadlalla A, Hindi NM. Is There a Place for Responsible Artificial Intelligence in Pandemics? A Tale of Two Countries. *Inf Syst Front*. 2021;
12. Timmermans S, Kaufman R. Technologies and Health Inequities. <https://doi.org/101146/annurev-soc-121919-054802>. 2020 Jul;46:583-602.
13. Hinton G, Bengio Y, LeCun Y. Recipients of ACM's A. M. Turing Award. Volúmenes 6-2019 de Communications of the ACM. Media A, editor. New York: Association for Computing Machinery; 2019.
14. OCDE. Recommendation of the Council on Artificial Intelligence (OECD Legal Instruments. OECD/ LEGAL/O449). France; 2019.
15. Miller T. Explanation in artificial intelligence: Insights from the social sciences. *Artif Intell*. 2019 Feb;267:1-38.

16. Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, *et al.* Deep Learning for Health Informatics. *IEEE J Biomed Heal Informatics*. 2017 Jan;21(1):4-21.
17. Bini SA. Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? *J Arthroplasty*. 2018 Aug;33(8):2358-61.
18. Darcy AM, Louie AK, Roberts LW. Machine Learning and the Profession of Medicine. *JAMA*. 2016 Feb;315(6):551-2.
19. Karakulah G, Dicle O, Kosaner O, Suner A, Birant CC, Berber T, *et al.* Computer Based Extraction of Phenotypic Features of Human Congenital Anomalies from the Digital Literature with Natural Language Processing Techniques. *Stud Health Technol Inform*. 2014;205:570-4.
20. Exploring how to prepare the healthcare workforce, through education and training, to deliver the digital future. London; 2019.
21. Turvey CL, Lindsay JA, Chasco EE, Klein DM, Fuhrmeister LA, Dindo LN. Current Practices in Electronic Capture of Patient Reported Outcomes for Measurement-Based Care and the use of Patient Portals to Support Behavioral Health. *Psychiatr Clin North Am*. 2019 Dec;42(4):635.
22. Tazawa Y, Liang K ching, Yoshimura M, Kitazawa M, Kaise Y, Takamiya A, *et al.* Evaluating depression with multimodal wristband-type wearable device: screening and assessing patient severity utilizing machine-learning. *Heliyon*. 2020 Feb;6(2):e03274.
23. Thorstad R, Wolff P. Predicting future mental illness from social media: A big-data approach. *Behav Res Methods*. 2019 Aug;51(4):1586-600.
24. Bernau C, Riester M, Boulesteix AL, Parmigiani G, Huttenhower C, Waldron L, *et al.* Cross-study validation for the assessment of prediction algorithms. *Bioinformatics*. 2014 Jun;30(12).
25. Sessa M, Khan AR, Liang D, Andersen M, Kulahci M. Artificial Intelligence in Pharmacoepidemiology: A Systematic Review. Part 1-Overview of Knowledge Discovery Techniques in Artificial Intelligence. *Front Pharmacol*. 2020 Jul;11:1028.
26. Lucas GM, Gratch J, King A, Morency LP. It's only a computer: Virtual humans increase willingness to disclose. *Comput Human Behav*. 2014 Aug;37:94-100.
27. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile Apps for Bipolar Disorder: A Systematic Review of Features and Content Quality. *J Med Internet Res* 2015;17(8)e198 <https://www.jmir.org/2015/8/e198>. 2015 Aug;17(8):e4581.
28. Jones N, Moffitt M. Ethical guidelines for mobile app development within health and mental health fields. *Prof Psychol Res Pract*. 2016 Apr;47(2):155-62.
29. Padrez KA, Ungar L, Schwartz HA, Smith RJ, Hill S, Antanavicius T, *et al.* Linking social media and medical record data: a study of adults presenting to an academic, urban emergency department. *BMJ Qual Saf*. 2016 Jun;25(6):414-23.

30. Zidaru T, Morrow EM, Stockley R. Ensuring patient and public involvement in the transition to AI-assisted mental health care: A systematic scoping review and agenda for design justice. *Heal Expect*. 2021;
31. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, *et al*. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA*. 2011 Aug;306(8):848-55.
32. Freeman K, Geppert J, Stinton C, Todkill D, Johnson S, Clarke A, *et al*. Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy. *BMJ*. 2021 Sep 1;n1872.
33. Lin SY, Mahoney MR, Sinsky CA. Ten Ways Artificial Intelligence Will Transform Primary Care. *J Gen Intern Med*. 2019 Aug;34(8):1626-30.
34. Graham J. Artificial Intelligence, Machine Learning, And The FDA. *Forbes*. 2017;
35. McKinsey Global Institute. A FUTURE THAT WORKS: AUTOMATION, EMPLOYMENT, AND PRODUCTIVITY. London; 2017.
36. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Futur Healthc J*. 2019 Jun;6(2):94.
37. Halevy A, Norvig P, Pereira F. The unreasonable effectiveness of data. *IEEE Intell Syst*. 2009;24(2):8-12.
38. Wen D, Khan SM, Ji Xu A, Ibrahim H, Smith L, Caballero J, *et al*. Characteristics of publicly available skin cancer image datasets: a systematic review. *Lancet Digit Heal*. 2022 Jan;4(1):e64-74.
39. Mooney SJ, Pejaver V. Big Data in Public Health: Terminology, Machine Learning, and Privacy. <https://doi.org/101146/annurev-publhealth-040617-014208>. 2018 Apr;39:95-112.
40. Caliskan A, Bryson JJ, Narayanan A. Semantics derived automatically from language corpora contain human-like biases. *Science (80- )*. 2017 Apr;356(6334):183-6.
41. Bohensky MA, Jolley D, Sundararajan V, Evans S, Pilcher D V., Scott I, *et al*. Data linkage: a powerful research tool with potential problems. *BMC Health Serv Res*. 2010;10.
42. Saeed SA, Masters RMR. Disparities in Health Care and the Digital Divide. *Curr Psychiatry Rep*. 2021 Sep;23(9).
43. Kickbusch I, Piselli D, Agrawal A, Balicer R, Banner O, Adelhardt M, *et al*. The Lancet and Financial Times Commission on governing health futures 2030: growing up in a digital world. *Lancet*. 2021 Nov;398(10312):1727-76.