

Inteligencia artificial en salud. Retos éticos y científicos

FUNDACIÓ
VÍCTOR
GRÍFOLS
i LUCAS

Inteligencia artificial en salud. Retos éticos y científicos

Cuadernos de la Fundació Víctor Grífols i Lucas
Inteligencia artificial en salud. Retos éticos y científicos Nº 63 (2023)
Edita: Fundació Víctor Grífols i Lucas. c/ Jesús i Maria, 6 - 08022 Barcelona
fundacio.grifols@grifols.com www.fundaciogrifols.org
ISBN 978-84-09-50075-8 Depósito Legal: B 7651-2023

Cuadernos
de la Fundació **63**
Víctor Grífols i Lucas

SUMARIO

Pág.

Presentación

Pere Ibern 7

Las implicaciones sociales de la genómica potenciada por la inteligencia artificial

Harry Farmer 13

Aplicaciones de la inteligencia artificial en nuestro entorno

La transformación de la salud mediante uso de datos e inteligencia artificial: ámbitos de aplicación, retos científico-tecnológicos y claves para la adopción

Felip Miralles 32

Inteligencia artificial ética y su aplicación en el ámbito de la salud: sistemas aplicados a la COVID-19

Alicia de Manuel 46

La inteligencia artificial en salud, ¿es «diferente»?

Juliana Ribera 60

La roboética en salud

Antoni Baena 71

Robótica social: aspectos éticos y sociales

Miquel Domènech 83

Relación de autores 98

Títulos publicados 100

PRESENTACIÓN

La inteligencia artificial ya está entre nosotros

En julio de 2022, la Fundació Víctor Grífols i Lucas dedicó sus Conferencias Egozcue anuales a la *Inteligencia Artificial en salud: retos éticos y científicos*, y este cuaderno recoge su contenido.

Para comprender lo que Bill Gates ha llamado «la tecnología más revolucionaria desde hace décadas», había que probarla y dejar constancia de ello. Haciendo este ejercicio, reproduzco los párrafos que han sido elaborados por un algoritmo de inteligencia artificial al ser preguntado sobre el tema. Se trata del ChatGPT, conocido por todos, en su versión de 23 de marzo de 2023. Estas son sus respuestas:

La Inteligencia Artificial (IA) ya está siendo utilizada con éxito en muchos ámbitos de la medicina, y se prevé que su uso siga creciendo en los próximos años. Existen diversos ámbitos en los que la IA ya está demostrando su potencial en la medicina, como: (1) Diagnóstico: los sistemas de IA pueden analizar datos médicos, como imágenes médicas o historias clínicas de los pacientes, para ayudar a los profesionales de la salud a realizar diagnósticos más precisos y rápidos. (2) Tratamiento: la IA puede ser utilizada para ayudar a diseñar tratamientos personalizados para pacientes individuales basados en sus datos genómicos y médicos. (3) Seguimiento de los pacientes: la IA puede ayudar a monitorizar la salud de los pacientes y detectar problemas o síntomas que puede haber pasado por alto un profesional de la salud. (4) Descubrimiento de medicamentos: la IA puede ser utilizada para analizar grandes cantidades de datos para identificar nuevas dianas terapéuticas y diseñar nuevos medicamentos.

En general, se prevé que la IA transformará la medicina en muchos sentidos, ayudando a mejorar la precisión de los diagnósticos, la eficacia de los tratamientos y la calidad de la atención sanitaria. Sin embargo, también se deben considerar las preocupaciones éticas y legales relacionadas

con el uso de la IA en la medicina, como la privacidad de los pacientes y la responsabilidad de los profesionales de la salud en la toma de decisiones clínicas.

La Inteligencia Artificial (IA) plantea varios problemas bioéticos que deben abordarse:

1. *Privacidad y seguridad de los datos:* la IA implica la recopilación, el almacenamiento y el uso de datos masivos. Es importante garantizar la privacidad y la seguridad de estos datos para evitar el robo de identidad, el ciberataque y la vulneración de la privacidad.
2. *Discriminación:* la IA puede ser programada para tomar decisiones que pueden ser discriminatorias por razones de género, raza, edad, orientación sexual, etc. Es importante asegurar que la IA sea imparcial y que no reproduzca prejuicios existentes en la sociedad.
3. *Transparencia y explicabilidad:* la IA puede tomar decisiones sin dejar claro cómo se ha llegado a ellas. Esto puede provocar que sea difícil comprender por qué se toman determinadas decisiones y dificultar la responsabilidad en caso de error.
4. *Responsabilidad:* la IA puede tomar decisiones que tienen consecuencias importantes para las personas y la sociedad en general. Es importante establecer quién es responsable de las decisiones que toma la IA y cómo se puede responsabilizar a los desarrolladores o a los propietarios de esta tecnología.
5. *Autonomía:* la IA puede llegar a ser autónoma en la toma de decisiones, lo que puede ser un problema cuando las decisiones que toma afectan a la vida de las personas. En este sentido, es importante establecer límites a la IA y asegurar que siempre se dé una supervisión humana.
6. *Ética de los vehículos autónomos:* la IA está siendo utilizada para desarrollar vehículos autónomos. Es importante garantizar que estos vehículos se comporten de manera ética en situaciones límite y que no pongan en peligro la vida de las personas.

En resumen, la IA plantea varios problemas bioéticos que hay que abordar para garantizar que esta tecnología se desarrolle de manera ética y responsable.

Hay diversos criterios que se pueden tener en cuenta para conseguir que una Inteligencia Artificial (IA) sea explicable:

1. *Transparencia:* la IA debe ser transparente en su funcionamiento, de manera que se pueda entender cómo se toman las decisiones. Esto incluye la capacidad de mostrar los datos utilizados para entrenar el modelo, así como los algoritmos y las reglas que utiliza para tomar decisiones.
2. *Interpretabilidad:* la IA debe ser interpretada por las personas, de manera que se pueda comprender lo que está haciendo y por qué está tomando las decisiones que toma. Esto implica una interacción humana y la capacidad de proporcionar explicaciones claras y comprensibles para los usuarios.
3. *Consistencia:* la IA debe ser consistente en su comportamiento, de forma que pueda ser confiable y previsible en su conducta. Esto implica la capacidad de proporcionar respuestas coherentes a situaciones similares.
4. *Robustez:* la IA debe ser robusta, es decir, debe ser capaz de manejar situaciones inesperadas o excepcionales sin caer en errores graves. Esto incluye la capacidad de gestionar datos no previstos o situaciones no programadas.
5. *Accesibilidad:* la IA debe ser accesible para todos los usuarios, independientemente de su nivel de conocimiento tecnológico. Esto implica la capacidad de proporcionar una interfaz de usuario clara e intuitiva, así como la capacidad de dar apoyo en diferentes idiomas y formatos.

En resumen, para conseguir que una IA sea explicable, es importante que sea transparente, interpretativa, consistente, robusta y accesible para todos. Esto ayudará a garantizar que la IA sea ética y responsable en su funcionamiento.

El principio de explicabilidad, según Luciano Floridi, es uno de los principios bioéticos fundamentales que se refieren a la necesidad de asegurar la transparencia y la responsabilidad en las decisiones tomadas por las Tecnologías de la Información y la Comunicación (TIC) que afectan a las personas.

Este principio se basa en el reconocimiento de que las TIC pueden tener un impacto significativo en la vida de las personas, especialmente en relación a la recopilación, el almacenamiento y el uso de los datos personales. Por lo tanto, las personas tienen derecho a entender y obtener explicaciones sobre cómo las TIC toman decisiones que afectan a sus vidas.

El principio de explicabilidad implica que las decisiones tomadas por las TIC deben ser justificables y explicadas en términos claros y comprensibles para las personas afectadas. Esto implica la necesidad de garantizar la transparencia en el funcionamiento de los algoritmos y los modelos de las TIC, así como la transparencia en la recopilación y el uso de los datos personales.

Este principio bioético también se relaciona con otros principios fundamentales como el respeto por la privacidad, la seguridad y la libertad de las personas. En resumen, el principio de explicabilidad es una parte importante de los principios bioéticos que buscan garantizar que las TIC sean utilizadas de manera ética y responsable para mejorar la vida de las personas y de la sociedad en general.

A la vista de esta experiencia, a fecha de hoy no podemos llegar a ser conscientes del cambio que representa la IA, del mismo modo que no lo fuimos cuando surgieron los ordenadores o los teléfonos móviles. Es lo que se dice ahora una metatecnología porque conviviremos con ella en muchísimos ámbitos diferentes. Solo con la simple lectura de estos párrafos, alguien puede haber pensado que yo mismo los he escrito, pero no es así. Ahora bien, yo desconozco de dónde ha sacado la información (transparencia) el ChatGPT y cómo ha sido elaborada (explicabilidad). Evidentemente, en este caso no le hemos pedido ninguna toma de decisión y, por lo tanto, no aparecen otras implicaciones bioéticas más relevantes señaladas anteriormente.

Todas estas implicaciones éticas de la inteligencia artificial en el ámbito de la salud fueron objeto de debate en las *Conferencias Josep Egozcue* de 2022. Han pasado 8 meses, pero hay que decir que la velocidad de aplicación desde entonces se ha multiplicado enormemente. En el mes de noviembre era accesible la primera versión de ChatGPT, y cerca de 100 millones de personas la

han estado utilizando desde entonces. En el mes de marzo de 2023 ha surgido la nueva versión y ya hay múltiples alternativas. Nos encontramos en un momento de eclosión de la tecnología que ha llevado a cerca de 1.000 expertos a pedir una moratoria en su utilización para que se garanticen unos principios éticos elementales.

Este es, por tanto, un momento crítico donde hay que aplicar un esfuerzo adicional para comprender una realidad nueva que nos impactará a todos. El cuaderno que tenéis en las manos es tan solo un primer ejemplo que quiere contribuir al debate y comprensión del tema.

Pere Ibern

Centre de Recerca en Economia i Salut. Universitat Pompeu Fabra
Miembro del Patronato de la Fundació Víctor Grífols i Lucas

F U N D A C I O N
V I C T O R I A
Q U E E N S
I N T E L L E C T U A L

Las implicaciones sociales de la genómica potenciada por la inteligencia artificial

Harry Farmer

Investigador del Instituto Ada Lovelace de Londres

Las implicaciones sociales de la genómica potenciada por la inteligencia artificial

La inteligencia artificial (IA) y la genómica están llamadas a convertirse en dos de las tecnologías más transformadoras del siglo XXI.

En los últimos años, estas dos tecnologías han ido imbricándose; de hecho, muchos avances en genómica han sido posibles gracias al despliegue de la IA, mientras que los datos genómicos se utilizan cada vez más y en mayores cantidades para entrenar a los sistemas de IA.

La convergencia de la IA y la genómica tendrá hondas repercusiones en la investigación médica, la asistencia sanitaria y las sociedades de todo el planeta, pero también plantea infinidad de cuestiones de orden jurídico-político.

Por sí solas, tanto la IA como la genómica suscitan interrogantes profundos sobre agencia humana, intimidad, igualdad, sesgos y poder. Cuando las dos tecnologías se analizan juntas, estos interrogantes se tornan muchísimo más complejos, teniendo en cuenta las múltiples maneras en que pueden evolucionar e interactuar, tanto entre sí como con factores externos.

Para los profesionales y gestores preocupados por cómo cultivamos, gestionamos y regulamos la ciencia de la genómica, la irrupción de la IA plantea, por tanto, toda una serie de preguntas importantes y difíciles de responder: ¿Dónde, cómo y hasta qué punto están cambiando las capacidades, las aplicaciones viables y el ejercicio de la ciencia genómica a raíz de la IA? ¿Qué cambios se prevén para el futuro? ¿Con qué certeza van a producirse esos cambios? Y, por último, ¿qué efectos sociales, políticos y económicos tendrán?

El objetivo de este artículo, adaptado de una conferencia impartida en junio de 2022, es doble. En primer lugar, se presenta un proyecto de investigación que estamos llevando a cabo para interrogarnos sistemáticamente cómo se está aplicando la IA y en concreto cómo se prevé aplicar a la ciencia genómica, además de explorar las ramificaciones éticas, políticas y jurídicas de las nuevas capacidades de la genómica.

En segundo lugar, el artículo describe el mundo que podría surgir como consecuencia del desarrollo y la proliferación de las capacidades genómicas potenciadas por la IA. No se pretende insinuar, con este segundo ejercicio, que la genómica potenciada por la IA termine configurando inevitablemente

un panorama injusto o desafortunado, ni siquiera que este pronóstico sea probable. Lo que se pretende es, por el contrario, ilustrar cómo determinados incentivos económicos podrían derivar en la adopción de tecnologías que, a pesar de su enorme potencial de beneficiar a la humanidad, se utilicen de forma contraria a los principios bioéticos básicos como los de autonomía, equidad, dignidad y solidaridad.

«AI and Genomics Futures»: un proyecto para investigar las repercusiones sociales de la genómica potenciada por la IA

«AI and Genomics Futures» es un proyecto conjunto del Instituto Ada Lovelace y el Nuffield Council on Bioethics pensado para investigar metódicamente las cuestiones éticas relacionadas con la convergencia de la IA y la genómica.

En el proyecto se utilizan diversos métodos –investigación, imaginación de futuros y participación pública– para explorar cómo la IA está transformando las capacidades y el ejercicio de la ciencia genómica, y qué significará esa transformación para las personas y para la sociedad.

«AI and Genomics Futures» se propone investigar dos grandes cuestiones sobre el desarrollo, la proliferación y el impacto de las capacidades genómicas que posibilita la IA.

En primer lugar, el proyecto se interesa por entender cómo los avances en inteligencia artificial (sobre todo aprendizaje automático y aprendizaje profundo) están modificando –y modificarán– las capacidades, las aplicaciones y el ejercicio de la ciencia genómica. El objetivo de este epígrafe es adquirir un conocimiento más claro de lo siguiente:

- qué puede hacer la genómica gracias a la IA, o qué puede hacer con mayor eficacia, y qué capacidades nuevas o mejoradas tendrá gracias a la IA;
- cuáles son las aplicaciones actuales de estas nuevas capacidades, cuáles están despuntando y cuáles cabe prever en el futuro;

Las implicaciones sociales de la genómica potenciada por la inteligencia artificial

- cómo está cambiando la forma de hacer genómica por la creciente relevancia de la IA, si es que está cambiando, así como los datos y la experiencia necesarios y los actores implicados.

En segundo lugar, el proyecto pretende explorar los posibles efectos sociopolíticos de las tendencias, las capacidades y las aplicaciones más relevantes de la genómica potenciada por la IA. Con un horizonte de entre 5-10 años, el proyecto explorará las repercusiones de las nuevas capacidades para secuenciar, analizar y comprender el genoma humano, así como su efecto sobre:

- los diferentes colectivos dentro de la sociedad y la distribución de los riesgos y beneficios de estas tecnologías;
- el reparto actual y futuro del poder económico y político;
- la prestación de asistencia sanitaria; la frontera entre la sanidad y otros ámbitos de la vida; y la relación entre el ciudadano, el Estado y el sector privado en la sanidad y en otros servicios públicos.

También indagará el efecto de otros factores extrínsecos inciertos, como las normativas de protección de datos, la titularidad de las técnicas genómicas y los avances y actitudes públicas en relación con la edición genómica.

Ámbito de estudio

El proyecto «AI and Genomics Futures» estudiará cómo la aplicación de la IA en genómica está afectando al análisis del genoma humano.

En este sentido, no se prestará atención al efecto de la IA en la edición genómica (por ejemplo, CRISPR) ni al análisis de genomas no humanos, salvo cuando haya una conexión evidente con el análisis de genomas humanos.

La prioridad del proyecto es, pues, explorar el posible impacto de los avances en genómica propiciados por la IA, concretamente en la capacidad de analizar el genoma o los genomas humanos.

Ello incluye avances que puedan repercutir, a su vez, sobre el análisis genómico o sus aplicaciones, como las mejoras facilitadas por la IA en secuenciación genómica, en el procesamiento de datos fenotípicos y en la comunicación y cesión de datos y conocimientos genómicos.

También se tendrán en cuenta los avances en el análisis del epigenoma humano (aunque con menor prioridad que los avances en análisis genómico).

Igualmente, los progresos en el análisis de genomas no humanos entran en el ámbito del estudio siempre que profundicen nuestro conocimiento del genoma humano (o de determinados genomas humanos) y de su interacción con genomas no humanos. Por ejemplo, entraría el análisis genómico de un virus para ver cómo interactúa con los genotipos humanos, pero no para vigilancia epidemiológica.

Por último, el proyecto no se limita a los avances en ciencia genómica que guarden relación con la medicina. También nos interesa la aplicación de la genómica en contextos distintos del sanitario.

Metodología

En cuanto a su metodología, el proyecto «AI and Genomics Futures» se divide en tres fases principales. La primera es una fase de investigación en la que se describirán de forma empírica las tendencias actuales y futuras en genómica potenciada por la IA. Esta investigación se realiza mediante tres actividades simultáneas:

- Una revisión bibliográfica, para evaluar la literatura académica actual sobre la aplicación de la IA a la ciencia genómica, además de los debates éticos, jurídicos y políticos en relación con la genómica potenciada por la IA.
- Un análisis «cientiométrico», realizado con técnicas informáticas, para analizar grandes conjuntos de datos (como índices académicos y repositorios de patentes, así como información pública sobre financiación para investigación), a fin de catalogar las tendencias actuales e incipientes en genómica potenciada por la IA.
- Una prospección de horizontes, efectuada con la técnica Delphi, en la que se preguntará a un grupo de expertos, concretamente, cuáles creen que serán los avances científicos más importantes en la intersección de la IA y la genómica en los próximos dos lustros.

Las implicaciones sociales de la genómica potenciada por la inteligencia artificial

La segunda será una fase de cribado. En ella, los directores del proyecto y el comité asesor determinarán las capacidades y tendencias en las que habrá que concentrarse en la fase exploratoria final. La selección se hará teniendo en cuenta su posible repercusión sociopolítica y económica, así como la probabilidad de que se hagan realidad en el próximo decenio.

Por último, en la fase exploratoria, se estudiarán las repercusiones que podría tener la genómica potenciada por la IA en la sociedad en general, teniendo en cuenta una serie de factores exógenos. Asimismo, se procurará dilucidar cómo los gestores actuales pueden y deben influir sobre el despliegue de las nuevas capacidades. En esta fase del proyecto se prevé utilizar los métodos siguientes:

- Topografía de escenarios: se trata de una herramienta de «imaginación de futuros», utilizada por empresas y gestores políticos para comprender y prepararse ante un futuro incierto. En lugar de intentar ofrecer una única visión de cómo será el futuro, se perfilan los distintos mundos que podrían configurarse partiendo de diferentes combinaciones de factores externos difíciles de predecir.
- Deliberación pública: una técnica particular para involucrar a la ciudadanía, en profundidad, en la toma de decisiones y la formulación de políticas públicas; consiste en facilitar a miembros del público el tiempo y la información necesarios para que emitan un juicio ponderado sobre una cuestión concreta y bien definida en materia de gestión política. En el caso de «AI and Genomics Futures», se invitará al público a deliberar en torno a qué creen que deben hacer los gestores políticos ante los posibles mundos que pueden surgir con el desarrollo y la proliferación de las capacidades genómicas.

Además de estas dos técnicas, en la fase exploratoria del proyecto también participarán expertos, se revisará la legislación y normativa vigentes en genómica y se determinará hasta qué punto coinciden la legislación y normativa actuales con las expectativas y prioridades manifestadas por el público en el ejercicio de deliberación.

El sistema sanitario tras la genómica

La segunda parte de este artículo está dedicada a indagar cómo pueden afectar las capacidades genómicas propiciadas por la IA en la concepción que tienen las sociedades del norte global sobre la asistencia sanitaria.

Los párrafos siguientes, por tanto, repasan algunas de las suposiciones sobre los cambios que podría traer la IA en genómica, así como las presiones económicas y sociopolíticas que pueden acabar determinando cómo se despliegan dichas capacidades, si es que se despliegan. Como se ha dicho en la introducción, este ejercicio está pensado, por una parte, para cartografiar un posible mundo futuro y, por la otra, para describir las circunstancias en que una tecnología con gran potencial positivo puede acabar empleándose con fines contrarios a algunas de las concepciones del progreso humano.

Mejora de las capacidades genómicas gracias a la IA

La primera tarea de este ejercicio es determinar exactamente qué se espera que mejore o cambie en la ciencia genómica por la aplicación de la IA.

Las investigaciones preliminares de «AI and Genomics Futures» revelan tres vías por las que la IA está mejorando –y se prevé que mejore– las capacidades de la ciencia genómica, a saber:

- En primer lugar, la IA contribuye a la genómica haciendo que la recogida de datos sea más fácil, más rápida y más eficiente. En particular, se ha dicho que el procesamiento del lenguaje natural (NLP) acelerará la digitalización de historias clínicas, ayudando a generar los datos fenotípicos necesarios para los análisis genómicos. Además, la IA permite efectuar la secuenciación genómica con mayor rapidez y precisión.
- En segundo lugar, la IA permite analizar mejor el genoma humano. Se ha afirmado que la IA mejorará el filtrado de variantes y la predicción de efectos y comportará avances en farmacogenómica y genética poblacional. También se ha vaticinado que algunos modelos de

aprendizaje automático podrían mejorar la precisión de las puntuaciones de riesgo poligénico.

- En tercer lugar, la IA está haciendo que sea más fácil y más práctico trasladar las informaciones genómicas al ámbito clínico. Cabe destacar aquí que la IA (y especialmente el procesamiento del lenguaje natural) puede facilitar sistemas más eficaces de asistencia a la toma de decisiones clínicas, ayudando a los médicos no especialistas a tomar decisiones a partir del análisis genómico; además, puede servir para trasladar a la clínica los datos genómicos, ayudando al personal sanitario a comunicar las implicaciones a los pacientes.

Todavía tenemos motivos de peso para ser cautos con estas predicciones. Sigue habiendo, de hecho, importantes interrogantes en torno a la exactitud intrínseca de las predicciones genómicas y la representatividad de los conjuntos de datos genómicos (muy sesgados hacia los individuos de ascendencia europea), lo cual se refleja en las dudas sobre la exactitud diferencial y los sesgos del análisis y la predicción. A ello se suman los debates candentes sobre la ética de delegar decisiones clínicas en los sistemas de asistencia por IA, junto con las incertidumbres sobre el uso de estos sistemas para comunicar diagnósticos y pronósticos a los pacientes.

En todo caso, el ejercicio realizado aquí se sostiene necesariamente (y muy artificialmente) en la hipótesis de que será posible solucionar estos problemas éticos y empíricos. En concreto, se da por sentado que, con conjuntos de datos suficientemente grandes y representativos, podrán solventarse los problemas de exactitud y los sesgos de los análisis genómicos y que, además, los sistemas de IA se desarrollarán hasta el punto de considerarse guías fidedignos y traductores fieles de la información genómica.

Al apartar temporalmente estos legítimos interrogantes sobre la exactitud del análisis genómico (junto con la responsabilidad de delegar la interpretación y comunicación de los datos genómicos en los sistemas de IA), podemos plantearnos qué reportará para los sistemas sanitarios la incorporación de la IA en el análisis genómico en el que quizá sea el escenario tecnológico más optimista.

En este escenario, la genómica potenciada por la IA sería capaz de ofrecer a los sistemas sanitarios:

- un mejor conocimiento de los factores de riesgo genéticos individuales;
- un mejor conocimiento de las respuestas individuales a las intervenciones sanitarias;
- la capacidad de desarrollar y desplegar estos conocimientos de forma barata y rápida.

La pregunta que cabe hacerse, por tanto, es la siguiente: ¿Qué implicaría la disponibilidad inmediata de este conocimiento para la prestación de asistencia sanitaria?

Antes de responder, tenemos que detenernos en varios factores previos. Primero, debemos esbozar las características definitorias de los sistemas sanitarios en el norte global de hoy. Segundo, tenemos que formular unas hipótesis explícitas sobre las presiones y los incentivos que afectan a los sistemas sanitarios y la forma en que estos pueden responder a las nuevas posibilidades tecnológicas.

Invirtiendo el orden de estos dos grupos de hipótesis es posible hacer unas previsiones cabales sobre la adopción, el despliegue y la regulación de las nuevas tecnologías en las economías capitalistas. Estas previsiones son las siguientes:

- En primer lugar, podemos suponer que, de no mediar factores compensatorios muy potentes, las nuevas tecnologías no se destinan necesariamente al mejor uso que puede dárseles. Por el contrario, su despliegue se ajusta al orden económico, de la forma que menos violenta los intereses imperantes. Este argumento lo ha expuesto convincentemente el filósofo y estadista brasileño Roberto Unger.
- En segundo lugar, según sostienen economistas como Carlota Pérez y Mariana Mazzucato, los sistemas socioeconómicos tardan mucho en reorientarse para incorporar plenamente el potencial de las nuevas tecnologías.
- En tercer lugar, tal como plantea el famoso dilema de Collingridge, es difícil revertir las primeras aplicaciones de una tecnología una vez adoptadas, al menos mediante regulación e intervención estatal.

También resulta posible y necesario formular algunas hipótesis político-económicas:

- En primer lugar, damos por sentado que los sistemas sanitarios seguirán sufriendo presiones económicas a corto y medio plazo.
- En segundo lugar, suponemos que gran parte de la capacidad en IA y genómica será desarrollada por el sector privado y permanecerá en sus manos, aun en el supuesto de que la financien y desplieguen los Estados.
- En tercer lugar, aventuramos que la tensión entre explotación y protección de datos personales se resolverá en favor de la explotación (al menos en el caso de la asistencia sanitaria).

Por último, es necesario hacer algunas hipótesis explícitas sobre las condiciones ambientales y demográficas que enmarcarán este ejercicio prospectivo:

- Damos por sentado que el envejecimiento tensionará los sistemas sanitarios y los servicios sociales.
- Damos por sentado que el calentamiento planetario agudizará paulatinamente la escasez de recursos y pondrá más presión sobre la salud de las poblaciones (sobre todo por los veranos cada vez más calurosos).

Además de estas hipótesis sobre los incentivos y las presiones ante los que se desarrollarán las capacidades de análisis genómico, también es preciso delinear las características esenciales de los sistemas sanitarios del norte global.

En cuanto a lo que ofrecen los sistemas sanitarios, cabe señalar que (pese a los prolongados esfuerzos por reconfigurarlos) siguen centrándose en tratar las enfermedades, en lugar de prevenirlas. También hay que mencionar que, en la actualidad, en la mayoría de los sistemas sanitarios del norte global, los recursos se asignan con criterios utilitarios, pero evaluando el efecto de la inversión en intervenciones, en vez de evaluar a quién se destinan dichas intervenciones.

Respecto a lo que esperan y requieren los sistemas sanitarios de los ciudadanos y pacientes, es notable que, en la actualidad, no se piden muchos datos a

la persona. Para quienes reciben tratamiento, la recogida de datos normalmente se limita a la información que resulta inmediatamente pertinente en la situación clínica. Además, las medidas de salud pública se adoptan predominantemente a nivel poblacional; son relativamente pocas las que se dirigen a grupos concretos.

Respecto a quién participa de la prestación, los Estados todavía tienen la capacidad de proporcionar la asistencia sanitaria con independencia de los actores privados, a pesar de la gran extensión de este sector.

Por último, los modelos sanitarios normalmente se financian y justifican siguiendo dos paradigmas.

En los modelos públicos, la prestación sanitaria normalmente se fundamenta en un principio de solidaridad. La justificación típica de la sanidad mantenida por los contribuyentes es que nadie sabe quién necesitará atención médica, de modo que el riesgo debe ser socializado.

Por el contrario, en los modelos de seguros, las aseguradoras trabajan con un paradigma de subvención recíproca, por el cual las personas que tienen pocas necesidades sanitarias acaban sufragando, a lo largo de su vida, la atención sanitaria de quienes la necesitan más.

Ambos modelos y sus justificaciones funcionan de forma distinta, pero los dos se basan, en cierta medida, en la idea de que los ciudadanos toman decisiones racionales sin saber cuánta atención sanitaria necesitarán en el futuro.

En este punto, hemos sentado todas las condiciones y suposiciones básicas para cartografiar nuestro mundo hipotético. Ahora podemos plantearnos cómo la disponibilidad inmediata de la genómica puede afectar al sistema. De nuevo, presuponemos que se habrán desarrollado las capacidades siguientes:

- una mejor comprensión de los factores de riesgo genéticos individuales;
- un mejor conocimiento de las respuestas individuales a las intervenciones sanitarias;
- la capacidad de desarrollar y desplegar estos conocimientos de forma barata y rápida.

Las implicaciones sociales de la genómica potenciada por la inteligencia artificial

Prevención de precisión

A pesar de los esfuerzos que vienen haciéndose desde hace mucho tiempo, hay grandes dificultades para implantar un modelo sanitario eminentemente preventivo. Cabe recordar que la prevención de las enfermedades crónicas características del norte global exige modificaciones ambientales. En particular, mejorar la vida de las personas no solo en el dominio de la salud, una tarea que, además de difícil de por sí, es políticamente controvertida.

Por otro lado, al no estar dirigida a grupos concretos, la prevención puede ser exagerada e insuficiente al mismo tiempo. La prevención de enfermedades mediante cambios conductuales supone una imposición importante sobre personas sanas que probablemente pueden permitirse algunos hábitos poco saludables. Y, al contrario, las iniciativas de salud pública dirigidas a modificar los comportamientos probablemente sean insuficientes entre las personas cuyos hábitos personales representan un riesgo verdadero.

Ahora bien, si llega un momento en que la genómica permite conocer con exactitud los trastornos a los que una persona es más propensa –y conocer mejor cómo responderá a un estímulo, como los alimentos, el ejercicio, el estrés o la falta de sueño–, será mucho más factible modificar el entorno y la vida de esa persona para evitar que enferme. Gracias a estos conocimientos, la prevención podría ser mucho más dirigida, sin imponer cambios indiscriminados en la vida de las personas y en la economía. En la práctica, esto significaría que las iniciativas de salud pública fuesen muy personalizadas.

En circunstancias futuras en que haya gran cantidad de datos sanitarios pero pocos recursos, esta transformación, que podemos describir como la adopción de un modelo de prevención de precisión, deviene del todo verosímil.

Asignación de recursos individualizada

El segundo cambio que podrían propiciar las mejoras en análisis genómico concierne a las decisiones sobre asignación de los recursos sanitarios.

En la actualidad, las decisiones no se toman a nivel de los pacientes, sino que los medicamentos y las tecnologías se asignan evaluando su efectividad relativa para el paciente medio.

Sin embargo, si la genómica nos permite discernir qué pacientes responden a cada intervención, los profesionales sanitarios (y los sistemas) podrían decidir si prescribir o no un fármaco o un tratamiento, incluyendo en la ecuación la probabilidad de respuesta individual determinada por la genética.

Mayor intervención farmacoterapéutica

Es concebible que las mejoras en análisis genómico también comporten una mayor prescripción de fármacos para la población general.

Ahora se sabe que cada persona responde de manera distinta a los medicamentos y que puede ser difícil predecir los efectos secundarios y reacciones adversas.

También hay cierta aversión social a la prescripción masiva de fármacos para trastornos crónicos comunes (ya sean de tipo psíquico o somático). Como resultado, el umbral para la intervención farmacológica se ha elevado bastante: la gravedad del problema tratado con un fármaco debe ser muy notoria para superar el riesgo de efectos secundarios.

No obstante, si los avances en farmacogenómica permiten predecir con mayor exactitud la respuesta individual a los medicamentos, este cálculo de los riesgos podría variar. Si resulta posible descartar que una persona vaya a sufrir los efectos secundarios frecuentes de un medicamento, será más fácil justificar su prescripción para trastornos menos graves.

Presión para ceder los datos genómicos

La otra cara de la moneda, en un sistema sanitario más personalizado, es el incremento de la presión sobre los pacientes y colectivos a fin de que cedan sus datos genómicos.

Para hacer predicciones precisas habrá que contar con conjuntos de datos mucho más grandes que los actuales, de modo que se tendrían que recopilar muchos más datos de pacientes y ciudadanos. Es posible, en tales circunstancias, que este *quid pro quo* se vuelva más explícito. Así, quienes decidan no compartir sus datos pero se benefician de la medicina genómica serían tacha-

Las implicaciones sociales de la genómica potenciada por la inteligencia artificial

dos de «polizones» o aprovechados; también podría introducirse la obligación de ceder los datos para quienes disfruten de los recursos sanitarios.

Modificaciones conductuales más personalizadas

Otro rasgo de los sistemas sanitarios que busquen paliar la escasez con la información aportada por la genómica es la personalización de los cambios conductuales que se piden a los ciudadanos.

Hoy en día, las iniciativas de salud pública persiguen una modificación del comportamiento de todos: por ejemplo, gravamos con impuestos las bebidas azucaradas, el alcohol y el tabaco, lo cual recae sobre todos los consumidores por igual.

Sin embargo, si la ciencia genómica consigue predecir quién sufrirá los efectos adversos de un hábito tóxico o una determinada alimentación, y quién no, será más difícil justificar medidas universales.

En ese caso, los Gobiernos tendrán que abandonar las medidas de salud pública –algo incompatible con el compromiso por una sanidad preventiva– o bien tendrán que dirigir dichas medidas exclusivamente a las personas que las necesiten.

Quebranto de la justificación de la sanidad colectiva

Uno de los cambios más significativo, pero también más incierto, que podría comportar la disponibilidad de la genómica atañe a la justificación de la asistencia y los sistemas sanitarios, que constituyen grandes y carísimas empresas colectivas.

Como ya he mencionado, en la actualidad hay dos grandes modelos para justificar y financiar la sanidad. Los modelos públicos (o los de seguro obligatorio) se fundamentan en un principio de solidaridad. Que las personas no sepan si enfermarán –y, por tanto, si necesitarán los recursos sanitarios– justifica la socialización o colectivización del riesgo.

Los modelos privados, en cambio, funcionan con un principio de subvención mutua. Como nadie puede predecir qué atención sanitaria necesitará, los seguros los contratan personas que en realidad no los necesitan, de modo que sufragan los costes de quienes sí harán un uso intensivo de los recursos a lo largo de su vida. Aunque funcionan de forma distinta, los dos modelos se basan, en cierta medida, en la idea de que los ciudadanos toman decisiones racionales con un conocimiento limitado sobre la atención sanitaria que necesitarán en el futuro.

Si la genómica permite discernir quién necesitará los recursos sanitarios a lo largo de su vida, la presión sobre ambos modelos aumentará de forma significativa.

En el modelo social, los ciudadanos contribuyen de buen grado porque no pueden descartar que ellos mismos o sus seres queridos tengan que recurrir al sistema en algún momento. El que todos ignoremos nuestra salud futura despierta un sentimiento de solidaridad.

Ahora bien, si los análisis genómicos permiten predecir, con un grado razonable de precisión, quiénes de nosotros serán beneficiarios netos del sistema y quiénes serán contribuyentes netos, es posible que este sentimiento de solidaridad se tambalee. El efecto sería especialmente agudo en condiciones de escasez persistente.

La desaparición de ese sentimiento de solidaridad podría comportar la fijación de condiciones para los supuestos beneficiarios netos: en efecto, el arreglo solo resultará atractivo para los contribuyentes netos si los beneficiarios hacen todo lo posible para reducir la carga que representan para el sistema.

En un modelo de seguros privados, podemos vislumbrar la misma dinámica pero con un mecanismo distinto. Los seguros privados funcionan como mutualidad: dependen de las personas que realmente no necesitan el seguro. Ahora bien, si se generaliza el acceso a los análisis genómicos –y es posible saber quién necesitará la atención sanitaria y quién no a lo largo de su vida–, el sistema se viene abajo. Solo contratarían seguros quienes fueran a necesitarlos, lo cual dificulta mucho las cosas a las aseguradoras.

Las implicaciones sociales de la genómica potenciada por la inteligencia artificial

A grandes rasgos, en ese supuesto pueden pasar tres cosas: la primera, que las compañías aseguradoras subiesen los precios desmesuradamente, lo cual no sería políticamente sostenible a largo plazo. Otra posibilidad sería que las aseguradoras intentasen ocultar a los posibles clientes sus perspectivas de salud según la genómica. Para ello tendrían que monopolizar el acceso a los análisis, algo que probablemente resultaría muy difícil tanto a nivel legal como en la práctica.

Una tercera posibilidad sería que se tomase la decisión política de obligar a contratar los seguros a todos los ciudadanos, aunque no fuesen a necesitarlos nunca. Llegado ese punto, el sistema se toparía con los mismos problemas mencionados sobre el quebranto de la solidaridad en los modelos sociales: quienes no necesiten el seguro se resentirían de tener que sufragar la atención médica de los demás y exigirán fijar unas condiciones a cambio de su aportación monetaria.

Conclusión: evaluar el posible mundo venidero

En su conjunto, los cambios y transiciones explorados en los párrafos anteriores prefiguran un mundo en el cual:

- la atención sanitaria es muy cara y se destina muchísima energía a evitarla de entrada;
- los ciudadanos no tienen capacidad de acción sobre sus datos genómicos y, por el contrario, la cesión de datos genómicos es parte del *quid pro quo* sanitario;
- resulta más difícil justificar la sanidad pública, al desvanecerse el sentimiento de solidaridad ciudadana y resentirse los contribuyentes netos ante los beneficiarios netos;
- los sistemas sanitarios aplican reglas y exigencias paternalistas a los ciudadanos, pero estas se siguen de forma desigual; los sistemas sanitarios practican una vigilancia constante de la salud de sus ciudadanos y los que tienen «genes de alto riesgo» son sometidos a una mayor vigilancia y mayores exigencias conductuales;
- los ciudadanos con genes de bajo riesgo aceptan sufragar la atención sanitaria de los demás únicamente si estos hacen todo lo posible por minimizar sus factores de riesgo genómicos;

- las farmacoterapias masivas se justifican más fácilmente que antes porque los efectos secundarios se pueden predecir mejor.

La pregunta que suscita este escenario es sencilla: ¿Es posible, en estas circunstancias, aplicar los principios bioéticos –agencia, solidaridad, intimidad, autonomía, dignidad, equidad– a nivel micro? Y si no, ¿qué pueden hacer los gestores políticos, los legisladores y los profesionales para garantizar que los sistemas sanitarios cobren una forma distinta cuando incorporen los análisis genómicos?

Esta es una de las grandes preguntas que nos haremos en el Instituto Ada Lovelace y el Nuffield Council, a lo largo del proyecto «AI and Genomics Futures», en relación con dicho mundo hipotético, así como con los muchos otros futuribles que nos depara la genómica potenciada por la IA.

**Aplicaciones de la
inteligencia artificial
en nuestro entorno**

La transformación de la salud mediante uso de datos e inteligencia artificial: ámbitos de aplicación, retos científico-tecnológicos y claves para la adopción

Felip Miralles

Director de Tecnologías de Salud, Eurecat

Técnicas de inteligencia artificial para la transformación de la salud

La Inteligencia Artificial (IA) o «Artificial Intelligence (AI)» es una disciplina del ámbito de las ciencias de la computación que propone métodos y técnicas para el desarrollo de programas informáticos con la habilidad de aprender y razonar como hacemos los humanos. El aprendizaje automático o «Machine Learning (ML)» es la parte de la IA que permite desarrollar algoritmos que tienen habilidad de aprender sin ser programados explícitamente. Más recientemente, un subconjunto del ML bautizado como aprendizaje profundo o «Deep Learning (DL)», designa a las redes neuronales artificiales que se adaptan y aprenden de enormes cantidades de datos.

En la transformación de la salud hacia un nuevo paradigma centrado en el paciente y basado en la evidencia y el valor, el uso de los datos y la aplicación de técnicas de DL, ML y en general de IA, son facilitadores científicos y tecnológicos fundamentales.

Efectivamente, el primer objetivo de la transformación de la salud es llevar a su usuario final principal, el paciente, al centro del sistema, con el objeto de mejorar su calidad de vida y su experiencia. La sostenibilidad de los sistemas de salud frente al cambio demográfico, el envejecimiento de la población y el aumento de la cronicidad, la dependencia y la discapacidad; la preparación

frente a los nuevos retos derivados del cambio climático y su efecto multifactorial en la salud de la población, que se ha hecho tan patente en la reciente pandemia global; unos retos y otros, importantísimos y complejos, traen a primer plano los objetivos subsiguientes, a saber: avanzar hacia una aproximación a la salud y la medicina más eficaz y eficiente, proporcionando métodos y herramientas que permitan aplicar la evidencia en lugar de la intuición o el oficio, y optimizar el valor como ratio de resultados frente al coste.

La digitalización de la salud, y en concreto, como estadio superior, la aplicación de la ciencia de datos y la IA a la salud y la medicina, llegan en el momento oportuno para facilitar esta transformación. A partir del análisis de las necesidades no cubiertas de los usuarios finales de los servicios de salud, principalmente pacientes y ciudadanos, también de cuidadores informales, normalmente familiares, y por supuesto de todos los profesionales de la salud (médicos de familia y especialistas, enfermeras, fisioterapeutas, nutricionistas, psicólogos y un largo etcétera), así como de gestores, administradores, decisores de todos los niveles asistenciales, el codiseño y la reingeniería de procesos permiten redefinir los nuevos servicios asistenciales.

Los datos recogidos en cada fase del viaje del paciente son la materia prima principal de este nuevo paradigma. Los datos clínicos del paciente, que incluyen variables demográficas, diagnósticos, exploraciones en visitas, resultados de pruebas, medicación, etc. y también imagen médica (radiografías, resonancias, TAC, endoscopias, etc.) se almacenan en grandes repositorios digitales de datos (historias clínicas electrónicas-EHR y archivos de imágenes-PAC; los acrónimos en inglés son los que utilizamos para designarlos).

Además, y dependiendo del tipo de reto y solución, otros tipos de datos cobran importancia creciente: en concreto el estudio de los datos biológicos de los pacientes, como la genómica (unos veinte mil genes del genoma humano), o de otras «ómicas» de interés, como son la transcriptómica (alrededor de un millón de moléculas de RNA mensajero que se encuentran en la célula), la proteómica (unos diez millones de proteínas que constituyen la manifestación fenotípica del genoma que necesita la célula en un momento dado), y la metabolómica (unos cien mil metabolitos catalogados hasta ahora, que muestran el perfil metabólico del paciente), en orden de complejidad creciente.

Otro tipo de datos es el cajón de sastre de los datos «ambientales»: son los datos que provienen de sensores biométricos (por ejemplo, inerciales para medir la cantidad e intensidad de la actividad física); sensores fisiológicos (por ejemplo, variables fisiológicas como la presión sanguínea, la glucosa en sangre, o señales eléctricas de cardiografía, encefalografía, o miografía) que recogen información del propio paciente; y sensores que recopilan datos del ambiente que le rodea, como temperatura, humedad o contaminación ambiental, entre muchos otros posibles.

Pues bien, todos estos datos que se recogen y almacenan en bases de datos poblacionales o individuales (convenientemente anonimizados y preservados para cumplir con las regulaciones de privacidad, seguridad y propiedad de la información personal, que en el caso de los datos de salud son de exigencia máxima) se procesan mediante técnicas estadísticas, pero sobre todo de IA, y más específicamente de ML y DL, para crear modelos predictivos y, consiguientemente, sistemas de soporte a la decisión para el diagnóstico, pronóstico y gestión terapéutica. El diagrama de la página siguiente recoge los cuatro ejes de las cuatro características que conforman la transformación de la salud mediante tecnologías de datos e IA.

La transformación de la salud está fundamentada en una aproximación preventiva basada en la promoción de los cuatro pilares de la vida saludable, a saber: actividad física continuada y adecuada que nos aleje de los riesgos del sedentarismo; nutrición y dieta equilibrada en cantidad y variedad según la pirámide de alimentos para evitar sobrepeso y obesidad; higiene del sueño para un descanso suficiente y reparador; y estado cognitivo y emocional saludable, con fuertes vínculos sociales, y evitando y ayudando a gestionar situaciones de estrés, tristeza o soledad.

Este modelo 4x4 centrado en el usuario, basado en la evidencia que proporciona la analítica de datos y la IA, y con una aproximación preventiva de vida activa y saludable, dará como resultado el despliegue de soluciones de medicina 4P: preventiva, para mantener a la población sana más tiempo y con mejor calidad de vida; predictiva, para avanzar al desarrollo de patologías, a la irrupción de episodios agudos, a las agudizaciones en enfermedades crónicas, de manera que se pueda actuar e intervenir antes de que sea demasiado tarde o de que el remedio sea más costoso en tiempo, dinero y sufrimiento para el paciente; personalizada, para especializar las intervenciones a las necesidades y características específicas de cada paciente; y participativa, permitiendo y animando al paciente a convertirse en un actor activo y proactivo del equipo de cura en todas las fases del ciclo de vida asistencial.



Figura 1. Modelo 4x4 de la transformación de la salud mediante ciencia de datos e IA

Desde el punto de vista más tecnológico, la aplicación de técnicas de IA requiere de la puesta en marcha y despliegue de infraestructuras de datos en las que implementar canales para el almacenamiento, preprocesamiento, procesamiento, validación y visualización de datos e información útil para todos los actores, especialmente pacientes y profesionales de la salud. Durante el preprocesamiento se realizan tareas de anonimización, limpieza, armonización, selección de características y minería de datos. La fase de procesamiento es en la que se aplican técnicas clásicas de IA (por ejemplo, basadas en reglas), de ML con aprendizaje supervisado (por ejemplo, árboles de decisión), de ML con aprendizaje no supervisado (por ejemplo, clusterización) o de DL (por ejemplo, redes neuronales artificiales) para construir los modelos que predicen, clasifican o razonan para proporcionar soporte a la decisión clínica basada en la evidencia y posibilitar las soluciones de medicina 4P.

Ámbitos de aplicación de la IA para la transformación de la salud

Clasificamos también en cuatro grupos los ámbitos de aplicación de la IA para la transformación de la salud, ámbitos que no son en absoluto estancos sino totalmente complementarios, ya que algunas soluciones complejas toman características a caballo de dos o más grupos:

1. Hospital inteligente del futuro u Hospital 4.0

La gestión de los hospitales y grandes centros de salud requieren de centros de control, modelos de gemelos digitales y paneles de control análogos a los que ya se vienen aplicando en los nuevos modelos industriales de la Industria 4.0. Estos nuevos modelos de gestión consisten en la recogida de datos sistemática en todos los ámbitos, departamentos y espacios del hospital mediante el despliegue de sensores y actuadores del Internet de las Cosas, que alimentan un centro de control que simula y mimetiza el funcionamiento en tiempo real del hospital, proporcionando la información y las alertas necesarias para la óptima toma de decisiones.

La aplicación de la ciencia de datos y la IA proporciona soporte inteligente a la planificación, optimización y reingeniería de los procesos asistenciales. También permite sistematizar el despliegue de nuevos procesos asistenciales basados en la atención domiciliaria y el hospital «líquido», de manera que la transición hospital-domicilio y domicilio-hospital para pacientes crónicos, pero incluso también para pacientes críticos y de paliativos, permita mantener a los pacientes de largo recorrido en su casa, donde mejor están normalmente, atendidos de manera continua y con el mismo rigor y control que si estuvieran ingresados; permitiendo, también, la hiperespecialización de los hospitales en aquellas tareas de máxima complejidad que requieren de un espacio hospitalario sofisticado en conocimiento y tecnologías, como por ejemplo la cirugía especializada.

La IA, complementada por otras tecnologías avanzadas como la robótica, la sensorial o la biotecnología, hace posible la automatización y el soporte inteligente a todos los servicios del hospital inteligente del futuro. Los dispositivos médicos cada vez más potentes, a la vez que efectivos, ubicuos y fáciles de utilizar, hacen posible la estandarización de diagnósticos, pronósticos, intervenciones y seguimiento.

Un ejemplo es el dispositivo médico Ephion Mobility de la empresa Ephion Health (<https://ephion.health/>), que proporciona un biomarcador para la valoración funcional de pacientes que sufren alguna patología que afecta la movilidad, mediante la recogida ubicua y en continuo de datos de sensores biomecánicos durante el protocolo de medida, y la aplicación de técnicas de IA para fusionar esos datos y proporcionar al profesional clínico una medida más objetiva del estado funcional del paciente y de esa manera mejorar el pronóstico y la personalización de los tratamientos.

2. Atención integrada para la gestión de cronicidad, discapacidad y dependencia

La práctica de la salud virtual permite reducir visitas presenciales y desplazamientos innecesarios, de manera que las visitas presenciales cara a cara de valor añadido se pueden priorizar y optimizar para que aporten el máximo valor a pacientes, familias, y también a los profesionales de la salud.

Además, los modelos de atención integrada permiten coordinar la atención al paciente entre los distintos niveles asistenciales, es decir entre la atención médica primaria, la atención médica especializada y la atención social. Dependiendo del estadio en que se encuentre el paciente en su viaje asistencial, la atención puede estar pivotando en la atención especializada hospitalaria, en episodios agudos como una cirugía mayor o una intervención compleja como, por ejemplo, un tratamiento oncológico; o pivotar en la atención primaria cuando el paciente está en los estadios de seguimiento regular del viaje asistencial, por ejemplo monitorizado a domicilio y con teleconsultas o visitas al ambulatorio cuando fuera necesario; o pivotando en la asistencia social cuando, por ejemplo, la intervención sobre el entorno familiar, social o económico de la persona tiene especial relevancia.

La aproximación de atención integrada requiere de integración de datos y procesos: la coordinación asistencial se hace posible si los distintos niveles asistenciales pueden consultar y actualizar información y tareas y compartirlas de manera transparente entre todos los niveles, también con el paciente. La recogida y compartición sistemática de datos y procesos tiene valor añadido si se aprovecha el dato para extraer conocimiento y aplicar técnicas de IA, ML y DL a fin de proveer de soporte inteligente a pacientes, cuidadores y profesionales de distinto tipo en los distintos niveles.

Tal soporte inteligente es un valor añadido de las herramientas mHealth de autogestión para pacientes y cuidadores, que proporcionan servicios de comunicación, monitorización, formación y seguimiento del proceso asistencial. Son servicios avanzados mediante la personalización y el soporte inteligente que proporcionan alertas y recordatorios basados en datos, las recomendaciones que promueven el cambio personalizado de comportamiento, por ejemplo para ayudar a mejorar la adherencia a tratamientos farmacológicos y no farmacológicos a partir del perfil específico de pacientes.

De manera análoga, el soporte inteligente es un valor añadido de las herramientas colaborativas para todo el equipo de profesionales, proporcionando soporte inteligente a la planificación y el seguimiento de la atención sanitaria y social, de manera que se introduzcan características de adaptación, flexibilidad y verdadera personalización.

La ciencia de datos y la IA facilitan el despliegue de modelos de atención integrada, claramente beneficiosos para el paciente y el resto de los actores, pero difíciles de aplicar en todo su potencial partiendo de una realidad de silos de información y procesos distintos según sea en cada momento la puerta de entrada del paciente a los sistemas de salud.

Un caso de uso de este tipo de soluciones es eVisió, en donde una herramienta de mHealth para llevar las pruebas oftalmológicas a domicilio del paciente integran una historia clínica oftalmológica que permite la colaboración de profesionales de primaria, oftalmólogos, optometristas y el propio paciente a lo largo de su viaje asistencial, en este caso en cribaje, diagnóstico y seguimiento de condiciones oftalmológicas.

3. Sistemas de soporte a la decisión clínica para una medicina basada en la evidencia

El análisis de datos en medicina, es decir la recolección, preprocesado y fusión de datos clínicos, epidemiológicos, biológicos y ambientales, así como la aplicación de técnicas estadísticas, especialmente técnicas de IA, ML y DL, permite crear modelos predictivos que se avancen al futuro, de manera que posibiliten predecir riesgos, morbididades y mortalidades en la detección y progresión de enfermedades. Un paso más allá es la creación de modelos prescriptivos que sugieran acciones en escenarios diferentes, asociadas a problemas de optimización, en los que algoritmos más sofisticados encuentren la solución óptima.

Los modelos predictivos y prescriptivos son el corazón de los sistemas de soporte a la decisión clínica para el cribaje, diagnóstico precoz, pronóstico y gestión terapéutica. Como se construyen a partir del análisis de grandes cantidades de datos, y aprenden de multitud de casos, asisten a médicos y otros profesionales de la salud a tomar decisiones mejor informadas, basadas en la evidencia, más precisas, reduciendo errores médicos y promoviendo la estandarización y la equidad, ya que un mismo sistema de soporte a la decisión desplegado en distintas áreas geográficas y utilizado por distintos profesionales con distintos niveles de especialización unifica y acerca la calidad asistencial a toda la población.

Los primeros casos de éxito en este ámbito están siendo los sistemas de soporte al diagnóstico y pronóstico basados en imagen médica radiológica, mediante la aplicación de técnicas de visión por computador y de ML o DL, de manera que el algoritmo aprende de una gran cantidad de imágenes de casos pasados, debidamente etiquetadas, para ayudar a identificar anomalías, determinar riesgos asociados a la patología y predecir su evolución. Estas herramientas les son de gran utilidad a los profesionales en la toma de decisiones, pero a menudo se pueden utilizar también por los profesionales junto a sus pacientes y familiares en escenarios de toma de decisiones compartidas, cuando la solución permite simular distintos escenarios futuros a partir de decisiones presentes, de manera que se pueden presentar al paciente distintas alternativas y así encontrar la más conveniente para ese caso específico. De nuevo, este tipo de soluciones médicas se benefician especialmente del progreso espectacular de la ciencia de datos y la IA en los últimos años.

Un ejemplo es DeepLung, un sistema que ayuda a los radiólogos a identificar, caracterizar y delimitar nódulos en el tejido pulmonar y, consiguientemente, mejorar el diagnóstico y pronóstico del cáncer de pulmón, proporcionando herramientas para predecir crecimiento y riesgo de malignidad del tumor en estadios tempranos de la enfermedad.

4. Soluciones de medicina y nutrición personalizadas

A partir del análisis de los datos biológicos de una persona, si somos capaces de modelar su perfil «ómico», podremos personalizar cualquier intervención de salud a sus características a fin de proporcionarle exactamente aquello que necesita. Esta es la promesa de la medicina personalizada, que utiliza el perfil biológico de un individuo para guiar las decisiones tomadas en relación con la prevención, diagnóstico y tratamiento de la enfermedad. El conocimiento del perfil biológico de un paciente puede ayudar a los médicos en la selección de la terapia más adecuada, así como administrar la dosis o la dieta apropiadas.

Es importante que la aproximación al perfil biológico sea holístico: no solo estamos hablando del genoma humano, del perfil puramente genético, sino

que en muchas aplicaciones este puede ser complementado o suplementado por las otras «ómicas»: transcriptómica, proteómica, metabolómica. Por ejemplo, el metabotipo de una persona puede ser determinante cuando hablamos de la personalización de una dieta nutricional.

En cualquier caso, de lo que se trata es de la comprensión y la interpretación de los mecanismos complejos de la salud y la patología desde una perspectiva biológica profunda, para ser capaces de adaptar y personalizar los tratamientos que se proponen, tanto los farmacológicos como los no farmacológicos, como por ejemplo las terapias de rehabilitación.

Por lo tanto, en este cuarto y último ámbito de aplicación, la ciencia de datos y la aplicación de distintas técnicas de IA abren la perspectiva de la personalización de la medicina (e intervenciones relacionadas, como la nutrición) para cada individuo, basada en la biología y la interacción con el fenotipo y los estilos de vida.

Es el caso de Preventomics (<https://preventomics.eu/>), un sistema que permite diseñar dietas nutricionales personalizadas y de precisión a partir del metabotipo y estilo de vida de cada individuo; la nutrición personalizada y de precisión se abre paso para proporcionar estrategias de prevención de enfermedades.

Retos científico-tecnológicos y claves para la adopción

1. Retos científico-tecnológicos y confiabilidad de la IA

Si bien somos capaces de identificar el potencial de la ciencia de datos y la aplicación de la IA para la transformación de la salud, así como clasificar los ámbitos de aplicación con ejemplos relevantes de investigación y desarrollo en cada uno de ellos, la realidad es que todavía cuesta destacar casos de éxito en los que se haya llegado a la adopción de soluciones efectivas en la práctica clínica real y la explotación sostenible en el mercado de las tecnologías médicas.

Las barreras a superar para esta adopción son múltiples, y están relacionadas con retos científico-tecnológicos que a menudo impactan en mayor o menor grado en factores clínicos y tecnológicos, pero también éticos y de negocio, a saber: asegurar la protección y privacidad de los datos; construir algoritmos con resultados que alcancen niveles de acierto y mitigación de errores suficientes; ofrecer esos resultados de manera que los usuarios finales los puedan comprender y utilizar de manera suficientemente útil y segura; que se cumplan todos los criterios éticos y que consiguientemente puedan pasar los criterios regulatorios que ya se están estableciendo en el tiempo; aplicar modelos de negocio que permitan asegurar la sostenibilidad de las soluciones, tanto para las empresas proveedoras que los comercializan y los mantienen como para los clientes proveedores de salud que los consumen; etc.

La solución a estos retos no tiene una receta sencilla. Ahora bien, una premisa necesaria es el cumplimiento de un marco legal para la IA confiable, que se define así: la IA confiable significa mitigar los riesgos para la salud de las personas, la seguridad y los principios fundamentales, así como respetar los valores éticos tales como la autonomía, la prevención del daño, la equidad y la explicabilidad de los algoritmos resultantes basados en datos e IA.

La Unión Europea viene desarrollando un marco legal para la IA confiable. En 2018 se empezó a aplicar la regulación GDPR (General Data Protection Regulation) que toda solución en salud basada en datos e IA debe cumplir, y que asegura la privacidad y el buen uso de los datos.

Recientemente, en abril de 2021, la Comisión Europea publicó la propuesta Artificial Intelligence Act: un marco legal completo para la regulación de la IA con el objeto de crear un ecosistema de la IA confiable. Para dar respuesta a este marco regulatorio, el desarrollo de soluciones de salud digital basadas en datos e IA deben utilizar una metodología de IA confiable por diseño, durante todo el ciclo de vida del codiseño, desarrollo, validación e implementación de soluciones, a fin de asegurar los preceptos siguientes:

- Calidad del proceso: adecuada evaluación de riesgos y medidas de mitigación.
- Calidad de los datos: óptima calidad de los datos que alimentan los sistemas para minimizar riesgos y resultados discriminatorios.

- Transparencia: información completa sobre el rendimiento y características de los sistemas; auditoría de actividades para asegurar la trazabilidad de los resultados.
- Explicabilidad: herramientas para ayudar a comprender e interpretar las predicciones de los sistemas, así como su propósito.
- Usabilidad: resultados e información clara y adecuada al usuario.
- Supervisión: medidas adecuadas de supervisión humana para minimizar riesgos.
- Fiabilidad: óptimo nivel de robustez, seguridad y precisión.

2. Claves para la adopción

Solo recientemente, en efecto, se han empezado a consolidar los retos, los marcos regulatorios y las metodologías para la adopción de la ciencia de datos y la IA para la transformación de la salud. Todavía hay pocos casos de éxito de adopción en la práctica clínica y el mercado, pero como hemos revisado en el presente artículo, hay grandes expectativas y esperanzas puestas en el potencial transformador de la tecnología.

En el marco del CIDAI (Centre of Innovation for Data tech and Artificial Intelligence de Catalunya), se encargó a un conjunto importante y heterogéneo de expertos la elaboración de un libro blanco sobre la IA para dar respuesta a los retos de los sistemas de salud (<https://cidai.eu/white-papers/>).

Este libro blanco identificó tres fases para la implementación ordenada y exitosa de una estrategia de adopción de la IA en las organizaciones de salud: en la primera fase, a corto plazo, las soluciones de IA a implementar son las más accesibles, en concreto las relacionadas con el apoyo a ciertas tareas manuales, administrativas y repetitivas. En la segunda fase, a medio plazo, las aplicaciones de IA se enfocan a empoderar a los pacientes, de manera que les permita acceder a soluciones desde su hogar que les monitorean su estado de salud y testeen sus síntomas. Y en una tercera fase, más a largo plazo, se espera que el estado de madurez de la IA en salud permita acceder a grandes bases de datos de estudios clínicos que faciliten detectar y predecir episodios en el desarrollo e interacción de las enfermedades y sus tratamientos.

Como recomendaciones claves para la adopción de las soluciones de datos e IA en salud, el libro blanco distingue entre las que pueden tener un gran impacto y viabilidad más a corto plazo, de las que puedan tener un impacto más a medio o largo plazo.

Entre las primeras, se recomienda para empezar el facilitar el acceso a los datos de salud: la creación de repositorios distribuidos y accesibles a los datos de salud, con gobernanza y procedimientos que permitan la compartición y revocación ágiles de datos de pacientes, garantizando validez y veracidad en todo momento, y regulando los controles de acceso y los flujos de datos, así como su uso adecuado, cumpliendo con la regulatoria, en concreto la GDPR, pero con la flexibilidad necesaria para potenciar la innovación. En segundo lugar, apoyar las estrategias de IA en las organizaciones de salud, partiendo del aprendizaje de las posibilidades y beneficios de su adopción, identificando oportunidades de mejora de los procesos a partir de necesidades no cubiertas, priorizando las acciones que maximicen el impacto minimizando los costes, con objeto de acelerar la incorporación de nuevos análisis de datos y modelos de IA que den respuesta a los retos identificados, para asegurar el éxito en el tiempo mediante la gestión sistemática utilizando buenas prácticas y estándares, y planificando la gestión del cambio, la formación de los trabajadores y la identificación de sinergias con otras organizaciones del ecosistema de salud. A continuación, en tercer lugar, la creación de entornos de pruebas para la validación y viabilidad de nuevos modelos de IA basados en datos, antes de aplicarlos a un ámbito tan crítico como el de la salud de las personas, en los que se prueben las herramientas mediante métricas y guías estándar antes de incorporarlas a entornos operativos reales. Y por último, en el paquete de recomendaciones a corto plazo, facilitar los partenariados público-privados, previa definición de los modelos de compartición y reparto de propiedad intelectual y económica ante la futura explotación de resultados, coordinando una estrategia común de innovación en salud entre los actores del ecosistema para maximizar el impacto y emplear muy bien los recursos, cumpliendo escrupulosamente con los marcos legales y éticos, y facilitando el acceso a financiación que cubra los recursos para una colaboración estable, combinando la estandarización de los procesos públicos de licitaciones con la obtención de inversión privada de fondos para la investigación e innovación en salud.

Entre las recomendaciones que miran más a medio largo plazo, la primera consiste en combatir la resistencia al cambio en las organizaciones, promoviendo el conocimiento entre los profesionales de la salud: por ejemplo, presentándoles casos de éxito sencillos, definiendo programas formativos en cultura de datos e IA adaptados a perfiles sanitarios, proveerles de dedicación horaria a la investigación e innovación, definiendo protocolos para la creación de equipos multidisciplinares de cocreación de soluciones IA, poniendo todo el énfasis en la transparencia de las mismas. A continuación, es fundamental la formación y retención de talento experto en el dominio altamente multidisciplinar de la ciencia de datos e IA junto con las ciencias de la vida y la salud, fortaleciendo el ecosistema local tanto tecnológico como de proveedores de salud, a fin de potenciar el emprendimiento en el territorio. Por último, en tercer lugar, pero no menos importante, está la sensibilización de la ciudadanía, distribuyendo información a pacientes y ciudadanos en general, diseminando la implementación de nuevos servicios y casos de uso en el sector salud, para minimizar la resistencia al cambio inicial asociado en términos de transparencia, ética, privacidad, seguridad e impacto en empleos, para favorecer la formación profesional continua y la gestión del cambio en las organizaciones.

Inteligencia artificial ética y su aplicación en el ámbito de la salud: sistemas aplicados a la COVID-19

Alicia de Manuel

Investigadora del Observatorio de Ética en Inteligencia Artificial de Catalunya

1. ¿Qué entendemos por inteligencia artificial ética?

Tradicionalmente, las definiciones de la inteligencia artificial (IA) se han encontrado asociadas a descripciones tecnológicas que tienen que ver con el uso de máquinas como ordenadores o robots que realizaban tareas que requieren de inteligencia humana. Sin embargo, actualmente comienza a existir una preocupación por definir la IA también a través de sus implicaciones, especialmente en cómo esta ha afectado a la sociedad.

Es importante comenzar este debate, ya que, si bien la IA puede tener un impacto positivo, ya sea haciendo que los procesos sean más seguros y eficientes, también puede llegar a suponer una amenaza, tanto en términos de seguridad de las personas como a través de la pérdida de bienes inmateriales como la privacidad, la intimidad, etc. Así pues, la IA ética pretende resolver o mitigar los problemas no deseables que se derivan de los riesgos de la implantación de los sistemas de toma de decisiones automatizados. De esta manera debemos señalar que construir sistemas de IA que adopten las normas éticas es esencial para generar un ecosistema donde la tecnología no perjudique a las personas, a otros seres vivos y sus hábitats.

Uno de los mayores retos para la implantación de una IA ética es la falta de un consenso global, ya que la percepción de la IA, su uso y aceptación está fuertemente influenciada por las normas culturales de los distintos países y regiones del mundo. Los modelos económicos y las distintas normas culturales, así como las características legales, tienen un fuerte impacto no solo en el estado

de la tecnología de IA sino también en el grado de adopción ético. En este sentido, actualmente podemos destacar tres modelos de IA que pugnan por convertirse en el modelo hegemónico en la carrera por el desarrollo de la IA y que constituyen tres maneras de entender esta tecnología:

- *AI for Profit* (inteligencia artificial para el beneficio) es un modelo establecido sobre todo en áreas como Estados Unidos y caracterizado por el liderazgo de un pequeño grupo de empresas que domina el desarrollo tecnológico y marca el crecimiento económico. Esto ha producido en los últimos años una serie de debates sobre el excesivo control que tienen un determinado grupo de entidades privadas y que definen los avances tecnológicos y el curso del mercado.
- *AI for Control* (inteligencia artificial para el control) es un modelo caracterizado por el uso de sistemas de IA como herramientas destinadas al control y a la seguridad. Podemos encontrar este modelo predominante en China, a través de ejemplos como la implantación del sistema de crédito social. Su uso de la IA está caracterizado por la centralización en un único órgano estatal de una gran cantidad de datos correlacionados.
- *AI for Society* (inteligencia artificial para la sociedad) es el modelo que pretende distanciarse de los dos modelos anteriores, en el cual priman tanto los derechos de los usuarios como los principios de una IA ética. Esta es la posición que está tomando la Unión Europea a través de diferentes iniciativas que tienen como objetivo la coordinación y regulación de distintos marcos legislativos. En este sentido, también debemos señalar que, aunque la adopción de los principios éticos de la IA pueda suponer un cambio enorme en la relación que establecemos con la tecnología, también puede suponer que la innovación tecnológica quede retrasada haciendo que otros mercados más competitivos y menos restrictivos aumenten su liderazgo.

2. ¿Cuáles son los principios éticos de la IA?

La primera referencia a la IA ética en Europa viene dada por el denominado High Level Expert Group in AI (HLEG AI), un grupo de expertos formado por la Comisión Europea que en 2019 comienza a sentar las bases de los prin-

cipios éticos de la IA a través del concepto de IA confiable (*trustworthy AI*). Según el documento titulado *Ethics Guidelines for Trustworthy AI*,¹ la IA confiable debe ser robusta desde el punto de vista técnico y respetar los principios y valores éticos teniendo en cuenta su entorno social.

Las directrices presentadas por el HLEG AI establecen siete puntos clave que los sistemas deben cumplir para poder ser considerados confiables:

- Agencia humana y supervisión: los sistemas deben empoderar a los seres humanos, así como garantizar los mecanismos de supervisión adecuados.
- Robustez técnica y seguridad: los sistemas deben ser seguros, precisos y fiables y asegurar planes alternativos en caso de error para minimizar y prevenir los daños no intencionales.
- Privacidad y gobernanza de datos: además de garantizar la privacidad y la protección de datos, los sistemas de IA deben tener en cuenta la calidad e integridad de los datos asegurando un acceso legitimado a ellos.
- Transparencia: los modelos de negocio de datos y sistemas de IA deben ser transparentes.
- Diversidad, no discriminación y equidad: en los sistemas de IA deben evitarse los prejuicios y sesgos injustos.
- Bienestar social y medioambiental: los sistemas de IA tienen que ser sostenibles y respetuosos con el medioambiente, ya que deben poder beneficiar a los seres vivos y las generaciones futuras.
- Rendición de cuentas: deben establecerse mecanismos que garanticen la auditabilidad y la rendición de cuentas tanto de los sistemas de IA como de sus resultados.

A partir de la constitución de estas bases el interés por determinar los principales principios éticos de la IA ha llevado a la publicación de numerosos trabajos, recomendaciones y marcos operativos. Un claro ejemplo de este interés es analizado por Anna Jobin en *The global landscape of AI ethics guidelines*,² que representa una actual falta de consenso entre los principios éticos de los documentos sobre IA revisados para el estudio. Dentro de la revisión de 84 documentos sobre principios éticos o guías para el desarrollo de sistemas de IA de

compañías privadas, agencias gubernamentales, partidos políticos e instituciones académicas, los principios más repetidos son: transparencia, justicia y equidad, no maleficencia, responsabilidad, privacidad, beneficencia, libertad y autonomía, confiabilidad, dignidad, sostenibilidad y solidaridad.

Esta falta de unidad nos indica dos cosas: por un lado, que existe un creciente interés por parte de organizaciones e instituciones de implementar una estrategia de transformación ética en los sistemas de IA en diferentes sectores, y, por otro lado, que algunas declaraciones vagas y poco específicas nos pueden hacer sospechar de la utilización de la ética como un lavado de imagen.

3. ¿Por qué es necesaria una IA ética?

Como podemos ver, en los últimos años la IA ética ha pasado de ser una cuestión del ámbito filosófico a convertirse en una necesidad tangible. El impacto de la tecnología en todos los sectores (industrial, cuidado de la salud, justicia, transporte, finanzas, entretenimiento...) y el aumento en la capacidad de procesamiento de datos, al mismo tiempo que la amenaza de una escalada armamentística con armas inteligentes ha generado un debate sobre aquellos principios que dan forma a una IA ética. Al respecto, Cathy O'Neil³ nos advierte de cómo el avance de modelos predictivos en general ha hecho que los sistemas de IA se conviertan en «armas de destrucción matemática» (*weapons of math destruction*), ya que si bien este tipo de tecnología se ha desarrollado como una herramienta que mejore nuestra calidad de vida, muchas veces su uso supone un elevado riesgo que puede llegar a producir resultados muy negativos a nivel social.

Es por eso por lo que a los beneficios en la utilización de sistemas de IA como pueden ser la optimización y automatización de procesos o la minimización de los errores, surgen también una serie de inconvenientes que va a ser fundamental tener en cuenta a la hora de valorar la idoneidad de aplicar la IA a un servicio.

También cabe destacar la necesidad de identificar uno de los más importantes riesgos de la utilización de la IA y que supone un problema ético a raíz de su utilización: los sesgos. Podemos definir «sesgo» como un peso desproporcio-

nado a favor o en contra de una cosa, una persona o un grupo en comparación con otro, generalmente de una manera que se considera injusta. Esto se produce porque, a medida que la IA se ha comenzado a implantar cada vez más en nuestra vida cotidiana, vemos cómo esta tecnología puede mantener o incluso amplificar diferentes sesgos negativos hacia distintos grupos de personas, como mujeres, personas mayores, personas con discapacidad o etnias minoritarias. Quizá uno de los casos más sonados sea el del algoritmo COMPAS utilizado para predecir el grado de reincidencia en presos en Estados Unidos. Esta herramienta de evaluación del riesgo delictivo se alimenta con datos históricos para hallar correlaciones entre factores como la edad o los antecedentes, y después utiliza correspondencias para predecir la probabilidad de reincidencia de un acusado. Sin embargo, la investigación realizada por ProPublica en 2016 reveló que esta herramienta estaba sesgada, pues las personas negras recibían el doble de probabilidades que las personas blancas de acabar en el grupo de alto riesgo.

Y ¿por qué se producen los sesgos? En general, se asume que cuantos más datos se utilicen para entrenar un modelo, mejor. Sin embargo, esto puede dar lugar a la utilización de datos poco representativos o directamente sesgados, de manera que el uso de bases de datos contaminadas o imprecisas dan lugar a resultados sesgados. Esto es lo que comúnmente se denomina *garbage in garbage out* (que podríamos entender como «basura dentro, basura fuera»).

Por otro lado, la utilización de datos que contienen desequilibrios implícitos o explícitos no solo refuerza una distorsión en los datos, también afecta cualquier toma de decisiones, haciendo que el sesgo pueda ser sistemático. Esto lo veremos más adelante, pero en el caso del empleo de sistemas de IA en situaciones de emergencia sanitaria como la que hubo durante la COVID-19 puede suponer que tu origen étnico marque la diferencia entre recibir o no un cierto tratamiento, en este caso un respirador.

Asimismo, nos podemos encontrar que el sistema de IA puede contener un sesgo debido a los propios sesgos implícitos o explícitos del desarrollador. Esto se produce debido a que gran parte del diseño de un determinado programa se basa en la comprensión del mundo del diseñador. De esta manera, es importante incluir en el proceso de desarrollo a diferentes agentes, usuarios y partes interesadas que vayan a estar afectados por el mismo.

Por último, la falta de comprensión en la manera que un sistema de IA ha llegado a un determinado resultado o en la manera en la que funciona supone un problema de transparencia y explicabilidad, motivo por el cual muchos sistemas de toma de decisiones algorítmicas son denominados actualmente «cajas negras» (*black box*, en inglés).

Este efecto impide conocer y validar el origen de una determinada respuesta, haciendo que puedan proliferar una gran cantidad de sesgos debido a que las correlaciones de los datos puedan estar produciéndolos. En este sentido, muchos autores proponen que la explicabilidad sea un criterio básico necesario para la toma de decisiones de la IA.

4. Sesgos en algoritmos de sistemas de IA desarrollados para la COVID-19

Si la implantación de los sistemas de IA ha poblado casi todos los ámbitos de la vida cotidiana, la lucha para paliar el avance y transmisión de la pandemia de la COVID-19 no iba a ser diferente. Es por eso que, a través del estudio que realizamos en *Bias in Algorithms of AI Systems Developed for COVID-19* (Delgado *et al.*, 2022)^{a4} analizamos el desarrollo de sistemas de IA tanto para triaje como para aplicaciones de monitorización y rastreo de contactos. Nuestro objetivo era entender los sesgos existentes en el uso de esta tecnología y los problemas éticos que podían aparecer:

Triaje

El triaje es un proceso de selección en el que los objetos se ordenan en categorías de acuerdo con criterios definidos por ciertas características, y es utilizado en el proceso de identificación de pacientes en situación de riesgo vital. De esta manera pudimos identificar una serie de sesgos:

a Esta *scoping review* forma parte del proyecto «Detección y eliminación de sesgos en algoritmos de triaje y localización para la COVID-19», dentro de la convocatoria «Ayudas Fundación BBVA a Equipos de Investigación Científica SARS-CoV-2 y COVID-19».

- La variabilidad de las fuentes de datos contribuyó a generar sesgos en el intercambio de las redes de investigación COVID-19 y esto afectó a la calidad de los datos. Autores como Sáez *et al.*⁵ analizaron cómo el hecho de mezclar datos provenientes de diferentes países de origen –en este caso la información de Filipinas provenía del Departamento de Salud, mientras que la de China procedía directamente de los informes de los pacientes– hizo que los resultados fueran inconsistentes, debido a la variación en la fuente de información.
- Encontramos sesgos en los modelos de predicción de la COVID-19 a causa de las muestras de datos no representativas.⁶
- El rápido desarrollo de los sistemas de IA conlleva un gran riesgo debido a los datos de entrenamiento sesgados y a la falta de reproducibilidad y de un recurso de datos COVID-19 regulado.⁷ Sin estrategias integrales de mitigación de sesgos, esto puede exacerbar las disparidades de salud existentes. «El código fuente de cualquier modelo de IA debe compartirse públicamente para garantizar que los modelos puedan aplicarse, generalizarse y compararse de manera transparente.»⁸

En cuanto a los problemas éticos en los algoritmos de IA utilizados para el triaje, sobre todo pudimos identificar riesgos en torno a la transparencia, que es esencial para comprender las predicciones y las poblaciones objetivo, los sesgos no reconocidos, los problemas de desequilibrio de clases y su capacidad para generalizar las tecnologías emergentes en entornos hospitalarios y poblaciones.⁹ Para garantizar que los modelos puedan aplicarse, generalizarse y compararse ampliamente, el código fuente de un sistema de IA debe compartirse públicamente y deben crearse marcos regulatorios para facilitar el intercambio de datos.

Aplicaciones de monitorización y rastreo de contactos

Las aplicaciones móviles de monitorización y rastreo de contactos están destinadas a determinar la proximidad entre una persona infectada y otras a través de dispositivos móviles utilizando las tecnologías wifi y Bluetooth. Durante el estudio pudimos identificar cuatro tipos:

- 1) rastreo de proximidad, que rastrea la cadena de transmisión de posibles contactos cercanos y movimientos para identificar quién pudo haber estado expuesto a una persona infectada;
- 2) respuesta al brote, que rastrea la gestión de los casos y sus contactos a través del ingreso de datos para evaluar las medidas de coordinación, mitigación y evolución de la enfermedad en la sociedad;
- 3) seguimiento de síntomas, que monitorea de forma rutinaria la evolución de los síntomas y signos autoinformados en un paciente para evaluar la prevalencia de la enfermedad y ayudar a informar de los procesos de rastreo de contactos;
- 4) aplicaciones de cumplimiento de cuarentena, que ayudan a los usuarios a realizar un seguimiento y control de sus propias cuarentenas.

En cuanto a los sesgos, pudimos identificar que se producían:

- Pérdida de datos o frecuencia de monitorización insuficiente: el desarrollo de aplicaciones sin control llegó a generar una recopilación de datos inadecuada y sesgos debido a la pérdida de algunos datos o a una frecuencia de monitoreo insuficiente, lo que puede conducir a la imposibilidad de comparar los datos recopilados de diferentes regiones.¹⁰ Aunque no afecta a la funcionalidad central de la aplicación, puede afectar al uso posterior de los datos recopilados: por ejemplo, la mayoría de los modelos de *machine learning* se han basado en datos provenientes de China, lo que puede llegar a limitar la escalabilidad a otras poblaciones.¹¹
- Los medios de información alteraron la naturaleza de las búsquedas, produciendo sesgos en áreas de potenciales conglomerados. Al respecto, identificamos un caso en Corea mediante el cual se monitorizaban los posibles infectados a través de las búsquedas que se realizaban en internet, de tal manera que cada vez que los medios informaban sobre la ubicación de un paciente con COVID-19 positivo, muchas personas que estaban cerca de la ubicación informada solicitaban información adicional relacionada con la COVID-19.¹²
- Estas aplicaciones presentan un alto riesgo de discriminación, especialmente para las personas afectadas (Mbunge, 2020).¹⁵ Específicamente, las aplicaciones de rastreo basadas en el Internet de las Cosas

(IoT) recopilan datos de toda la población en tiempo real que luego se analizan para mapear los puntos críticos de la COVID-19. Dichos datos incluyen información étnica, detalles demográficos y estado socioeconómico que pueden influir en la asignación y distribución de los recursos destinados a la COVID-19 y producir discriminación.

- Los falsos negativos se convirtieron en un obstáculo, ya que pueden generarse deliberadamente porque las personas infectadas no quieren revelar su verdadero estado, siendo un factor que genera datos imprecisos sobre la población.

En cuanto a los problemas éticos, pudimos identificar diferentes riesgos éticos:

- Por un lado, en cuanto a la *privacidad*. El uso de aplicaciones de monitorización y rastreo de contactos plantea preocupaciones éticas, legales, de seguridad y de privacidad (Roche, 2020). Para ser aceptable, esta injerencia en los derechos fundamentales debe ser justificada, razonable, proporcionada y políticamente consensuada. Las aplicaciones de monitorización y rastreo de contactos brindan poca o ninguna privacidad a las personas infectadas y les exigen que divulguen sus datos, lo que plantea problemas difíciles de consentimiento, privacidad, ética y compensaciones entre bienes públicos y privados. Por ejemplo: la correlación de datos, el intercambio de información y la capacidad de extraer información de diferentes puntos de entrada contribuyen a la creciente fragilidad de la anonimización de los datos. Esta anonimización es aún más frágil cuando la información se recopila a lo largo del tiempo y mediante referencias cruzadas de datos.¹³
- Por otro lado, identificamos una *falta de regulación* al no existir regulaciones específicas en torno a este tipo de tecnología. El uso de datos, acceso o privacidad se ha adaptado a las leyes internacionales, nacionales y estatales. Aunque las regulaciones de algunos países protegen mejor a los ciudadanos, quedan dudas sobre hasta qué punto los Gobiernos y las instituciones privadas tienen acceso ilimitado y a largo plazo de los datos de los ciudadanos con fines de vigilancia.¹⁴
- Por último, problemas relacionados con el *consentimiento*. Los usuarios tienen derecho a darse de baja y configurar sus dispositivos,

poniendo en peligro la monitorización de casos positivos.¹⁵ Las aplicaciones de monitorización y rastreo de contactos deberían permitir que las personas practiquen la retirada del consentimiento,¹⁶ ya que es posible que los usos problemáticos de las tecnologías continúen una vez que termine la pandemia, como todo lo que tendría que ver con los beneficios financieros y políticos al perpetuar el uso de las tecnologías. En este sentido obligar a una población a usar una aplicación puede conducir a tasas de cobertura mucho más bajas.¹⁷ Sin embargo, encontramos escenarios opuestos en países que no han desarrollado ninguna aplicación específica. Brasil, por ejemplo, ha aumentado su vigilancia tecnológica para minimizar la cadena de transmisión de la COVID-19.¹⁸ Esta aplicación de vigilancia masiva puede plantear problemas sobre el poder, el abuso y la explotación de datos.

5. Sistemas de autoevaluación ética de sistemas de IA para aplicaciones de salud

En los últimos años, como respuesta para paliar los problemas éticos de la implantación de sistemas de IA, se han desarrollado diferentes herramientas de autoevaluación ética para verificar la adecuación de un sistema a diferentes principios éticos. Han sido muchas las instituciones que han elaborado sus propias guías de adecuación, como por ejemplo la ya mencionada *Ethics Guidelines for Trustworthy AI*,¹⁹ *Draft text of the Recommendation on the Ethics of Artificial Intelligence*,²⁰ *Declaración de Barcelona*,²¹ *AI Ethics Impact Group: From Principles to Practice*,²² *Technical methods for regulatory inspection of algorithmic systems in social media platforms*²³ o *AI systems classification framework at the OECD*,²⁴ entre otras.

En este sentido, y como resultado de la identificación de sesgos y de problemas éticos del proyecto de «Detección y eliminación de sesgos en algoritmos de triaje y localización para la COVID-19», nos propusimos la ideación de una *Guía de autoevaluación ética de uso de IA en salud* con el objetivo de contribuir a una toma de conciencia de la importancia de detectar y eliminar la injusticia algorítmica tanto en el diseño como en la aplicación de los algoritmos usados en el campo de la salud. Es importante señalar que este

documento está orientado a todos los agentes involucrados en el diseño e implementación de sistemas de IA en el ámbito de la salud, aunque puede ser igualmente útil para cualquier persona interesada en la detección y eliminación de la injusticia algorítmica.

Esta guía se encuentra en un formato de formulario o listado de preguntas que evalúa el sistema en todas las fases de su ciclo de vida y contempla áreas como la justicia, la equidad y las dinámicas participativas, los errores de juicio, la adecuación con la práctica clínica, la transparencia y la privacidad.

Por otro lado, cabe destacar también el papel de otros modelos de autoevaluación que, si bien no son específicos para el área del cuidado de la salud, tienen una naturaleza holística aplicable a otros sectores como es el caso del Modelo PIO (Principios, Indicadores y Observables), una propuesta de autoevaluación organizativa sobre el uso ético de datos y sistemas de IA que ha sido desarrollado por el Observatori d'Ètica en Intel·ligència Artificial de Catalunya con el soporte del Departament de la Vicepresidència i de Polítiques Digitals i Territori de la Generalitat de Catalunya, la Universitat de Girona y la Fundació i2cat. Este modelo de autoevaluación en forma de *checklist* (disponible en www.oeiac.cat) está basado en una recopilación de la literatura publicada y evalúa todo el ciclo de vida de los sistemas de IA en base a siete principios éticos: la transparencia y explicabilidad, la justicia y equidad, la seguridad y no maleficencia, la responsabilidad y rendición de cuentas, la privacidad, la autonomía y la sostenibilidad.

6. Conclusiones

A medida que la implantación de sistemas de IA avanza hasta suponer un importante punto de inflexión en el apoyo a la toma de decisiones, sobre todo en el ámbito de la salud, es necesaria la concienciación de los agentes implicados no solo en cuanto al conocimiento de la herramienta, sino también en el impacto ético que puede suponer su uso. Las herramientas de autoevaluación ética son una buena aproximación a esta toma de conciencia, pero sin un ejercicio formativo y legislativo corremos el riesgo de que la inteligencia artificial ética acabe convirtiéndose en una etiqueta en vez de en una meta a conseguir.

Notas

1. Comisión Europea. Ethics Guidelines for Trustworthy AI. Independent High-Level Expert Group on Artificial Intelligence. 2019, 8 de abril. Disponible en: <https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf>
2. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. *Nat Mach Intell.* 2019; 1: 389-399. Disponible en: <https://doi.org/10.1038/s42256-019-0088-2>
3. O'Neil C. *Weapons of Math Destruction.* Londres: Allen Lane; 2016.
4. Delgado J, et al. Bias in algorithms of AI systems developed for COVID-19: A scoping review. *Journal of Bioethical Inquiry.* 2022; 19(3): 407-419. Disponible en: <https://doi.org/10.1007/s11673-022-10200-z>
5. Sáez C, Romero N, Conejero JA, García-Gómez JM. Potential limitations in COVID-19 machine learning due to data source variability: A case study in the nCov2019 dataset. *Journal of the American Medical Informatics Association, JAMIA.* 2021; 28(2): 360-364. Disponible en: <https://doi.org/10.1093/jamia/ocaa258>
6. Röösl E, Rice B, Hernandez-Boussard T. Bias at warp speed: how AI may contribute to the disparities gap in the time of COVID-19. *Journal of the American Medical Informatics Association, JAMIA.* 2021; 28(1): 190-192. Disponible en: <https://doi.org/10.1093/jamia/ocaa210>
7. Casiraghi E, et al. Explainable Machine Learning for Early Assessment of COVID-19 Risk Prediction in Emergency Departments. *IEEE access: practical innovations, open solutions.* 2020; 8: 196299-196325. Disponible en: <https://doi.org/10.1109/ACCESS.2020.3034032>
8. *Ibidem*, n. 7. p. 191.
9. *Ibidem*, n. 7.
10. Ravizza A, Sternini F, Molinari F, Santoro E, Cabitza F. A Proposal For COVID-19 Applications Enabling Extensive Epidemiological Studies. Pro-

- cedia Computer Science. 2021; 181: 589-596. Disponible en: <https://doi.org/10.1016/j.procs.2021.01.206>
11. Scott IA, Coiera EW. Can AI help in the fight against COVID-19? The Medical Journal of Australia. 2020; 213(10): 439-441.e2. Disponible en: <https://doi.org/10.5694/mja2.50821>
 12. Hisada S, et al. Surveillance of early stage COVID-19 clusters using search query logs and mobile device-based location information. Scientific Reports. 2020; 10(1): 18680. Disponible en: <https://doi.org/10.1038/s41598-020-75771-6>
 13. Roche S. Smile, you're being traced! Some thoughts about the ethical issues of digital contact tracing applications. J Locat Based Serv. 2020; 14: 71-91. Disponible en: <https://doi.org/10.1080/17489725.2020.1811409>
 14. Marabelli M, Vaast E, Li JL. Preventing the digital scars of COVID-19. European Journal of Information Systems. 2021; 30(2): 176-192. Disponible en: <https://doi.org/10.1080/0960085X.2020.1863752>
 15. Mbunge E. Integrating emerging technologies into COVID-19 contact tracing: Opportunities, challenges and pitfalls. Diabetes & Metabolic Syndrome. 2020; 14(6): 1.631-1.636. Disponible en: <https://doi.org/10.1016/j.dsx.2020.08.029>
 16. Ibidem, n. 15.
 17. Moretti A, Santi C. Commentary to «Critical Limitations of Digital Epidemiology: Why COVID-19 Apps Are Useless». Survey Research Methods. European Survey Research Association. 2020.
 18. Ibidem, n. 15.
 19. Ibidem, n. 1.
 20. UNESCO. Preliminary report on the first draft of the recommendation on the ethics of artificial intelligence. 2021. Disponible en: <https://unesdoc.unesco.org/ark:/48223/pf0000374266>
 21. IIIA CSIC Barcelona. Declaration for the Proper Development and Usage of Artificial Intelligence in Europe. 2017. Disponible en: <https://www.iiia.csic.es/barcelonadeclaration/>
 22. AIEI Group (AI Ethics Impact Group). From Principles to Practice: An interdisciplinary framework to operationalise AI ethics. 2020. Disponible en: <https://www.ai-ethics-impact.org/en>
 23. Ada Lovelace Institute. Technical methods for regulatory inspection of algorithmic systems in social media platforms. 2021, diciembre. Disponible en: <https://www.adalovelaceinstitute.org/report/technical-methods-regulatory-inspection/>
 24. OECD. OECD Framework for the Classification of AI systems. OECD Digital Economy Papers. 2022; 323. OECD Publishing, París. Disponible en: <https://doi.org/10.1787/cb6d9eca-en>

La inteligencia artificial en salud, ¿es «diferente»?

Juliana Ribera

Medical Consultant. Amalfi Analytics

El valor de una tecnología depende de su utilidad, de su contribución a mejorar algún aspecto de relevancia social, incorporando para ello todos los análisis de coste-beneficio que consideremos pertinentes. Llevar a cabo este análisis de forma rigurosa y transparente es, a nuestro entender, un compromiso ético irrenunciable en el sector de la salud y en cualquier otro ámbito de la actividad humana, muy especialmente en el sector público.

Por lo tanto, y en respuesta a la pregunta que da título a este artículo, la aplicación de la inteligencia artificial (IA) a la gestión sanitaria se rige –o debería regirse– por los mismos valores que inspiran la evaluación de cualquier otra tecnología aplicada al gobierno de la *cosa pública* y de la misma manera que se valora cualquier otra tecnología en el ámbito de la sanidad.

La sanidad en general, y la sanidad pública en particular, es un sector especialmente sensible porque su razón de ser es procurar la buena salud de las personas y, por lo tanto, de la sociedad. Si bien es cierto que aquí es más crítico evitar un mal uso de la IA (y de cualquier otra tecnología), también lo es que en este sector es fundamentalmente crítico aportar todas las soluciones disponibles que puedan mejorar la gestión y aplicación de unos recursos limitados.

La IA no es más que una tecnología de la información que nos ofrece la posibilidad de, en primer lugar, vislumbrar patrones que, de otra manera, quedarían ocultos y, en segundo lugar, predecir resultados que difícilmente podríamos haber acertado. ¿Es esto relevante para la sanidad pública? ¿Aporta algún valor para la gestión eficaz y eficiente de los recursos públicos que destinamos a ella?

Nosotros creemos que sí, y lo argumentamos a continuación.

La desconfianza

En el entorno sanitario las reticencias de profesionales y pacientes se hacen más patentes; no obstante, la inteligencia artificial ya está dentro del sistema sanitario y su penetración no deja de aumentar.

Tendemos a considerar que el entorno de la salud es «diferente» a otros ámbitos y que el tratamiento de la información no puede ser igual que en ellos y tiene que ser «especial».

Efectivamente, la relación entre salud y bienestar la sitúa como un dominio político, con valores como la equidad, la accesibilidad y otros que no están presentes (o no tanto) en otros sectores comerciales, financieros e industriales donde la IA se utiliza con pocas restricciones. Las preocupaciones éticas se deben especialmente a la reducción de la supervisión de los expertos en determinados procesos de aprendizaje automático.

Esta desconfianza se ve aumentada por la responsabilidad de la salud individual y la salud pública y los determinantes sociales de la salud. La consideración de la salud como un servicio público hace que se espere de la Administración un sistema activo en la prevención, más que en la respuesta a problemas ya patentes. Es aquí donde los modelos predictivos pueden ser de más ayuda.

Podemos poner como ejemplo el uso de la analítica predictiva con *big data* por parte del National Center on Homelessness among Veterans,¹ que anticipa el riesgo de acabar en una situación de sin techo, y así evitar casos.

Balancear los beneficios y las consideraciones de respeto a datos personales y sensibles es obligado si queremos garantizar la salud, y no solo curar la enfermedad. La tecnología IA se evidencia como cada vez más crítica para responder a las necesidades de la ciudadanía. Por esto, desde la Administración se pone como condición poder defender el despliegue ético de estas tecnologías.

Parece más complejo, en efecto, que en otros sectores; sin embargo, empresas con un uso más intensivo de IA están invirtiendo en servicios de salud, y en principio con el mismo tratamiento.

Amazon ha adquirido recientemente empresas para prestar atención primaria, atención domiciliaria y servicios de farmacia en línea. Bajo la marca Amazon Care entra en el sector de la salud con fuerza.²

Sin embargo, Google Health cambió su estrategia en 2021 por la dificultad de aplicar las mismas pautas en salud que en otros sectores.

Entre otros motivos, debemos considerar que los procesos asistenciales tienen que ver más con el contexto y el modelo que con la tecnología en sí. La relación de confianza entre el médico y el paciente es lo diferencial y para proteger esta confianza el marco legal no puede ser el mismo que en otros servicios.

En todos los países de la Unión Europea se aplica la base del Reglamento General de Protección de Datos, que presume de ser uno de los más restrictivos en el uso de los datos de salud. Supone unas obligaciones en el tratamiento y unas recomendaciones de protección. Los conceptos de *protección de datos por defecto* y *desde el diseño*³ son clave a la hora de plantearse innovaciones en este campo.

Considerar el respeto a la privacidad y la protección de datos sensibles desde el principio de un desarrollo supone utilizar procedimientos de encriptación, maximizar la anonimización, realizar los correspondientes análisis de riesgo y análisis de impacto, e implementar otras actuaciones que en ocasiones suponen una restricción al potencial de la tecnología.

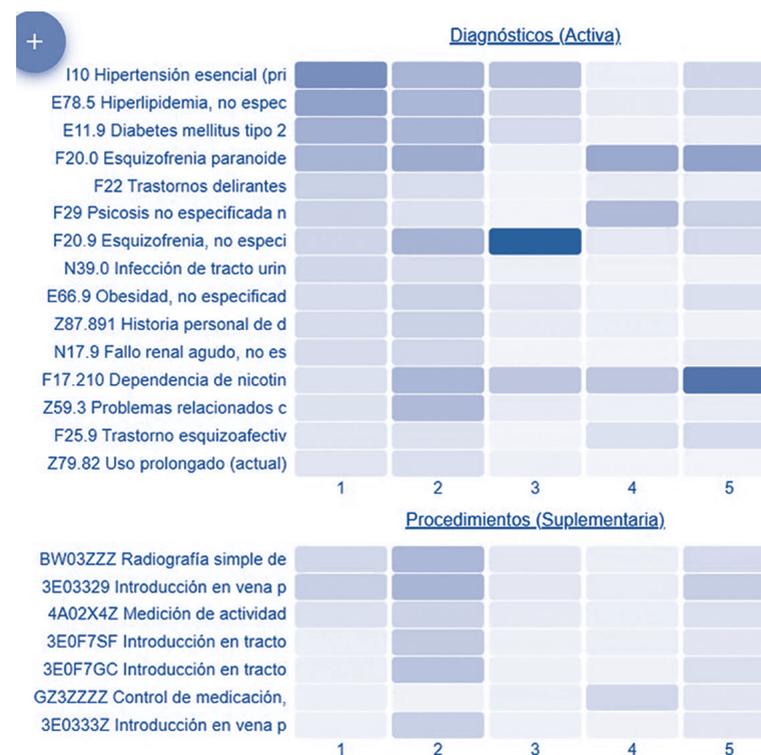
También aplicar protocolos que garantizan la trazabilidad del procesamiento de datos, su seguridad ante accesos inapropiados, medidas de notificación y unas normas restrictivas de temporalidad y acceso a los datos.

Sin embargo, las situaciones que amenazan la vida y la economía hacen saltar estas restricciones. Hemos tenido el doloroso ejemplo de la pandemia de COVID-19 y hemos visto cómo la excepcionalidad ha permitido poder trabajar y avanzar en muy poco tiempo con el análisis de los datos clínicos, que en circunstancias normales no se podrían procesar sin un largo período de trámites de consentimientos.

Un ejemplo lo tenemos en el programa COVID-X. Para abordar los diferentes retos, se ha dispuesto de una plataforma de datos que han permitido aportar

soluciones en poco tiempo. Ante una situación desconocida, los algoritmos de aprendizaje de máquina, o aprendizaje no supervisado, permiten a los expertos analizar los diferentes estados de salud con que los pacientes entran en el sistema, su evolución intrahospitalaria y los componentes diferenciales entre estados con una mejora evolutiva y estados que comportan riesgos. Este es un estudio, de los muchos realizados, que ha permitido un mayor conocimiento del proceso en el Hospital de la Santa Creu i Sant Pau.⁴

Se demuestra así que es posible mantener unos niveles correctos de garantías, desburocratizando los procesos de investigación y facilitando el progreso. Esperemos que no se retroceda a las restricciones extremas y limitantes para avanzar en las mejoras asistenciales.



Cluster	1	2	3	4	5
Número de pacientes	752 (28.36%)	362 (13.65%)	410 (15.46%)	873 (32.92%)	255 (9.62%)
Edad media	66.92	68.95	55.28	42.77	47.48
Sexo (Hombres, Mujeres)	51% 49%	55% 45%	62% 38%	59% 41%	74% 26%
Ratio de readmisiones a 7 días	12.79%	17.49%	7.89%	7.61%	18.16%
Ratio de readmisiones a 30 días	25.92%	38.34%	17.56%	20.22%	36.53%
Ratio de éxitos a dos años	-	-	-	-	-
Estancia media (días)	7.18	6.7	5.65	8.5	6.62
Alta en domicilio (continuidad asistencial por AP)	800 (66.89%)	540 (67%)	341 (61.11%)	778 (65.82%)	315 (65.76%)
Defunción	38 (3.18%)	31 (3.85%)	14 (2.51%)	5 (0.42%)	5 (1.04%)

Por supuesto, los niveles en los que se compromete la privacidad y la seguridad son diferentes según el nivel de aplicación de la tecnología IA.

Se debe tener un diferente tratamiento para investigación, asistencia, docencia y gestión.

La gran utilidad de la analítica predictiva a la hora de mostrar escenarios de futuro que faciliten las decisiones de programas territoriales de salud, reorganización y planificación del sistema de salud y la gestión de instituciones sanitarias no se está aprovechando.

En este nivel, no se aplican todas las restricciones legales a fin de tratar los datos de forma agregada garantizando la privacidad y la anonimidad. Las decisiones informadas en la utilización de recursos de manera más eficiente suponen un beneficio general para los pacientes, y no implican en ningún caso una

decisión opaca de un algoritmo sobre una persona. Las consideraciones éticas, aquí, siguen siendo importantes.

La entrada de la IA en los hospitales se ha priorizado en el punto de atención, con soporte a decisiones clínicas. Sin embargo, cada vez es más claro el gran beneficio de estas soluciones en los puntos de gestión de pacientes y recursos.

Es sumamente ético utilizar los escasos recursos sanitarios con la máxima eficiencia y donde más se necesitan, y además es una manera de introducir la IA en los centros de salud y crear una cultura de decisiones híbridas IA y humanas, que mejoran resultados.⁵

Siendo tan claro, ¿por qué no está utilizándose de forma masiva?

Las barreras que se observan para incorporar la IA en el día a día de la gestión sanitaria son múltiples, y entre ellas se suele citar la limitación económica. No obstante, mayoritariamente apuntan a la falta de formación en IA y a que ello hace mantener un arraigo en los cuadros de mando, que no ven la complementariedad y valor añadido de la información predictiva.

Una vez más, no es un proceso tecnológico, sino de liderazgo y gestión del cambio.

Otra barrera muy importante la encontramos en el abismo digital social. La población más anciana y con menos recursos está siendo excluida de los servicios digitales también en el ámbito de la salud. Este sí que es un tema que se debe abordar teniendo bien presente el componente ético del derecho universal a la salud.

En el documento sobre estrategia digital en el Reino Unido⁶ se debate la contradicción que supone la garantía de accesibilidad a todos los servicios con el programa de digitalización, que está excluyendo de estos servicios a unos nueve millones de ciudadanos. Esta situación es la que también vivimos, ya que se da en todos los países que han priorizado la digitalización.

La participación de la ciudadanía en la modelización del tratamiento de sus datos, así como la garantía de no excluir a parte de los ciudadanos de los servicios con IA, es crítica.

Muchas iniciativas para establecer un marco que minimice riesgos sociales y contemple los retos éticos de la IA deben acabar confluyendo en unas recomendaciones específicas para cada nivel de tratamiento.

Sobre conceptos claros de constituir un beneficio social, excelencia científica, evitar sesgos, y otros como podemos encontrar en el decálogo de uso de algoritmos de Google⁷ en el caso del tratamiento de datos de salud, debemos añadir varias consideraciones más, referidas especialmente a que respondan a necesidades reales de los pacientes, a que sean transparentes con los datos utilizados, a que se pueda disponer de una trazabilidad del uso de los datos, y a que el beneficio sea evidente bajo criterios científicos.

Se está abordando la elaboración de los principios de creación y uso de algoritmos de IA con principios éticos. Una base muy coherente la encontramos en la Declaración de Montreal, que apunta aspectos de participación democrática y solidaridad entre países y entre generaciones, un desarrollo sostenible y respetuoso con el entorno, y una organización independiente de ciudadanos que constituyan un observatorio del uso e impacto de la IA. Esta declaración también contiene una serie de recomendaciones muy específicamente enfocadas a los entornos tecnológicos de desarrollo de algoritmos.

Somos muy conscientes de la necesidad de disponer en las organizaciones de salud de una estrategia de gobernanza de datos para mejorar la integración de diferentes fuentes, su usabilidad, seguridad y cumplimiento de estándares. Al introducir programas con IA, la consistencia y confiabilidad de los datos se hace más crítica y el cumplimiento de las nuevas regulaciones ha de ser parte de los procedimientos y políticas del servicio.

La IA será omnipresente en los procesos del sistema de salud, en todos los niveles y, por lo tanto, se debe avanzar en el gobierno de datos, y plantear el modelo de *gobierno de los algoritmos*, para garantizar un buen uso.

Sin duda, hay factores diferenciales muy específicos para tratar los datos de salud, pero en realidad también en educación y otros ámbitos públicos se debe llegar a un mismo nivel de exigencia.

Lo cierto es que, a pesar de las prevenciones que ocasiona la IA, con el miedo a la sustitución de las personas, cada vez se entiende más que la IA ayuda a una mejor gestión sanitaria y a la humanización de la medicina.

El cardiólogo e investigador Eric Topol, en su libro *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*, explica con muchos casos prácticos, reales, como la IA contribuye a mejorar el diagnóstico y el cuidado de los pacientes, reduciendo la carga de trabajo de los facultativos.

Asimismo, reflexiona sobre el uso, en ocasiones poco preciso, del concepto IA. El término IA, en sí, podría estar creando una falsa percepción. Diversos expertos consideran que la IA no existe. En verdad, está en sus inicios, a pesar de sus más de treinta años de historia como campo de conocimiento y, sin entrar en el debate filosófico de qué es la inteligencia, podemos afirmar que los algoritmos y sistemas de cálculo que constituyen la IA, incluso el *machine learning*, no son «inteligentes».

Lo cierto es que no existe (todavía) y que tenemos la oportunidad de desarrollar una regulación que permita el avance científico, respetando a las personas. El avance es imparable. Muchas técnicas quirúrgicas ya no se conciben sin el soporte de la robótica y la IA. Cada vez más procesos tendrán este soporte, en el día a día, y en la consulta médica. Es un momento de expansión, y es el momento del equilibrio regulatorio, para no frenar, pero, a la vez, no poner en riesgo a las personas.

La introducción de nuevas tecnologías en la salud, sobre todo en ese espacio de la relación entre el paciente y la persona o las personas que cuidan de su salud y bienestar, siempre es motivo de incertidumbre, de preocupación, de un cierto miedo. Cuando se introdujo el ordenador en la consulta, muchas voces aseguraban que no se aceptaría jamás por parte del paciente. Sin embargo, hoy sería impensable la visita al médico sin la presencia del ordenador.

Esta tecnología no es como las anteriores en cuanto a su introducción y merece estos debates y esta amplia participación para obtener el máximo de su potencial. Pero el debate tiene que ser honesto y tenemos que evitar los discursos de miedo que tan a menudo encontramos en según qué medios. La IA merece ser debatida y juzgada en las mismas condiciones que otras tecnologías y bajo las mismas reglas éticas porque la salud merece tener la sostenibilidad y mejora que ofrecen las soluciones digitales y la IA.

Ninguna tecnología está exenta de riesgos –la IA tampoco–. Por eso, resulta imprescindible aplicar el principio de prudencia, y hacerlo al mismo tiempo que se promueve la investigación y los avances que permitan mejorar nuestro sistema de salud.

En conclusión, la pregunta que deberíamos hacernos es: siendo la sanidad un sector tan fundamental para el bienestar de las personas y para el progreso y la cohesión de la sociedad, ¿es ético no desarrollar y emplear herramientas a nuestro alcance –incluida la IA– que pueden contribuir a mejorar su gestión, que pueden ayudar a asignar mejor los limitados recursos, logrando así una mayor eficacia y eficiencia y, en consecuencia, unos mejores resultados en salud para el conjunto de la población?

Notas

1. <https://www.va.gov/homeless/nchav/index.asp>
2. https://cincodias.elpais.com/cincodias/2022/07/21/companias/1658429481_769311.html
3. https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/what-does-data-protection-design-and-default-mean_en
4. <https://www.santpau.es/web/public/-/sant-pau-i-amalfi-analytics-lideren-el-projecte-dinnovacio-traject>
5. <https://amalfianalytics.com/es/la-gestion-asistencial-como-mejor-via-de-entrada-de-la-inteligencia-artificial-en-los-hospitales-2/>
6. <https://www.gov.uk/government/publications/uks-digital-strategy/uk-digital-strategy>
7. <https://ai.google/principles/>

La roboética en salud

Antoni Baena

Director del Programa del Máster Universitario en Salud Digital (MUSD). Estudios de Ciencias de la Salud. Universitat Oberta de Catalunya (UOC)

Introducción

La pandemia de COVID-19 ha sido la primera pandemia que la sociedad ha tenido que afrontar, una pandemia que ha impactado sobremanera en muchos aspectos como la economía, la educación o las relaciones sociales, pero, por encima de todo, sobre la salud física y mental. De hecho, si se ha podido superar o rebajar dicha tensión ha sido gracias a la tecnología, que ha sido capaz en un tiempo récord de desarrollar diferentes vacunas, por una parte.

Por otra, si nos centramos en otros aspectos de la salud como el impacto que el aislamiento haya provocado, especialmente en algunos grupos vulnerables como personas mayores, dependientes o sin acceso a internet, la tecnología fue también una solución. Además de utilizarse para evitar el aislamiento y poder, de alguna forma, monitorizar el estado de nuestros seres queridos, también se aceleraron los diferentes procesos de tecnificación de la atención sanitaria a distancia, campo que llevaba un ritmo incluso demasiado lento para las necesidades y demandas de la sociedad. Gracias al uso de internet muchos pacientes pudieron seguir con sus tratamientos o incluso iniciar otros nuevos, ante la incapacidad de acceder inicialmente a un centro presencialmente. El final de la distancia física se ha aceptado como cotidiano e incluso como deseable o imparable. Primero como una simple sustitución o reemplazo de tecnologías: hacer lo mismo, pero con alguna mejoría (más rápido, más barato, más limpio...), pero lo mismo. Sin embargo, el gran salto proviene del uso de la tecnología para conseguir un proceso nuevo, diferente y transformador, es decir, innovador.

Nos hemos acostumbrado a utilizar la tecnología en cada ámbito de nuestra vida diaria (asistentes virtuales, chatbots, aplicaciones para todo, etc.), de forma que a veces no somos conscientes de estar hablando con una máquina, pues los sistemas de inteligencia artificial (IA) ya son capaces de tomar ciertas decisiones, aparentemente poco importantes, por nosotros. Lo que no deja de parecer una tarea intelectual, si no fuera porque la propia máquina no sabe el porqué de la decisión tomada. Pero ¿quién toma esa decisión, con base a qué, quién se hace responsable, se han tomado medidas de control, se ha pensado en posibles efectos iatrogénicos o sesgos de género, raza o ideología? En otras palabras, ¿se ha tenido en cuenta la ética o la deontología previamente a la implantación de la tecnología?

La tecnología no consiste en sustituir lo que ya tenemos, sino en reemplazar un instrumento o un procedimiento por herramientas digitales transformadoras, innovadoras; y es, precisamente, esta capacidad de innovación la que hace que muchas veces se haya dado prioridad a la capacidad tecnológica por encima de la beneficencia humana. Las máquinas, hasta hace relativamente poco, sustituían o facilitaban las tareas físicas, hasta que se fueron implementando soluciones que empezaron a asumir tareas intelectuales. Primero, las repetitivas y automatizadas, liberando a las personas de las tareas más complejas y, habitualmente, las de mayor importancia, calidad y productividad.

Esta evolución llevó al desarrollo de la IA, una disciplina que intenta replicar los procesos inteligentes mediante la computación. En definitiva, mediante la programación y análisis de datos la tecnología es capaz de crear la falsa ilusión de inteligencia, razonamiento y aprendizaje, palabras que hasta ese momento solo se relacionaban con las funciones cognitivas humanas.

Esta dependencia tecnológica, que en un inicio fue valorada muy positivamente, a medida que iba pasando el tiempo, provocó diferentes reacciones, siendo uno de los aprendizajes más claros el hecho de que el uso de la tecnología en la salud requería una mayor y mejor planificación, valoración y, sobre todo, de profesionales con competencias suficientes. Pero estábamos ante una emergencia sanitaria global, no había tiempo que perder.

Precisamente esa urgencia y aceleración hizo, en su momento, casi imposible un análisis objetivo, con lo que las críticas al uso de las tecnologías arriesgaron aún más. Es hoy cuando se está valorando qué aspectos son mejorables, cómo se hubiese desarrollado la pandemia con unos procesos tecnológicos analizados, probados y efectivos, qué se tenía que priorizar para que en situaciones parecidas la respuesta fuese mucho mejor.

La robótica

Dos aspectos de la salud digital han sido los más valorados. Uno el que hace referencia a las competencias digitales de los profesionales de la salud, ya que se ha constatado que el nivel de competencias digitales aplicadas era mucho menor del previsto, con lo que es muy necesario analizar y formar a todo el

colectivo sanitario. El segundo aspecto que ha adquirido importancia es la roboética: la ética aplicable a las máquinas.

Hemos llegado a un punto de desarrollo social en el que se piensa en un problema, se mira a la tecnología como solución y, por último y si se llega a hacer en algún momento, se valora si dicha solución es idónea u ocasiona más problemas de los que soluciona. Si la solución se implementa y, por casualidad o por azar, no provoca ninguna incidencia, se sigue con su uso sin más crítica. Solo se revisa si se detecta algún problema, no habiendo dado importancia a la prevención de dichas situaciones, como debería ser. Y esto sucede principalmente porque la ética o las buenas prácticas suelen darse por hecho y los profesionales implicados o no han sido formados en esta disciplina o no se les requiere. Lógicamente, a medida que la tecnología se ha ido adentrando en la salud, cada proceso es más complicado que el anterior y, por tanto, es más factible que se trate de procesos delicados y sensibles. Paralelamente, las personas sobre las que se aplica la tecnología también se han ido preocupando y preparando cada vez más para su autocuidado, tomando decisiones personales y empoderadas sobre su salud.

En definitiva, habitualmente se nos ha intentado convencer de que la tecnología es neutra, cuando es falso. Así nos podíamos olvidar o dejar en segundo plano la importancia de diseñar y controlar éticamente cualquier nueva aplicación tecnológica y, únicamente, centrarnos en los supuestos beneficios. Como es lógico, la ideología es diseñada y aplicada por personas con ideología, y esta se transmite a su producto.

La casi ausencia de planificación ética ha provocado una mayor dependencia con respecto a la tecnología ya que, en lugar de adaptar esta a las necesidades reales, se ha planificado pensando más en la adaptación del usuario a la tecnología. Aunque claramente es una situación en rápido retroceso, principalmente porque los usuarios y usuarias de salud han dado un paso al frente para controlar diferentes procesos y así evitar convertirse en simples espectadores de las actuaciones sobre su salud, incluso gracias a decisiones algorítmicas arbitrarias no fundamentadas, ni eficaces, ni claras. El negocio está en la predicción y sin datos no es posible; por eso, la captación de datos, más allá incluso de su calidad, es lo que mueve actualmente a muchos desarrolladores.

Este aumento de la preocupación y el interés por los derechos digitales quedó claro cuando en enero de 2022 se publicó la Declaración Europea de los Derechos y Principios Digitales, cuyo contenido establece que: 1) las personas son el núcleo de la transformación digital; 2) se debe apoyar la solidaridad y la inclusión; 3) garantizar la libertad de elección y expresión; 4) aumentar la seguridad y empoderamiento de los ciudadanos, y 5) promover la sostenibilidad en un entorno digital fiable, diverso, multilingüe y seguro. Es decir, el objetivo final no es ni más ni menos que intentar equiparar los derechos *offline* y *online*, aunque no dejan de ser unos principios más descriptivos que normativos, ya que no son vinculantes.

Pero previamente ya había habido algunas reuniones de expertos como la que se realizó en 2017 en Asilomar (California, EE. UU.), cuyo resultado fueron los Principios Asilomar de la Inteligencia Artificial, donde se consensuaron 23 principios que deben servir de guía para el análisis ético-deontológico de cualquier tecnología aplicada.

La importancia de los datos

Difícilmente se podrá innovar si no se tienen datos suficientes como para plantear escenarios y respuestas múltiples. Hasta ahora se le ha dado un valor excesivo a la captación de los datos, sin tener en cuenta su fiabilidad y su representatividad, como si el hecho de tener la capacidad de computar millones de datos superase la importancia de realizar una buena retroalimentación y valoración constante de los resultados. De hecho, en muchos casos las matemáticas se han confundido con la realidad, de forma que la realidad resultante solo existía teóricamente. No olvidemos que la IA proyecta un futuro analizando datos del pasado, confunde resultados con realidad, de forma que la realidad ya no importa, incluso no existe, solo los resultados marcan el camino.

La ética en los chatbots

¿Qué hay más humano y complejo que el lenguaje y la comunicación? Por eso, en la década de 1960, se empezaron a desarrollar aplicaciones que simulaban conversar interaccionando con respuestas automáticas previamente estableci-

das. El primer chatbot de la historia fue diseñado por el informático Joseph Weizenbaum en 1966 cuando construyó un programa que simulaba una conversación con una psicóloga llamada Eliza. Es decir, la primera simulación cognitiva de una máquina iba dirigida a resolver problemas cognitivos. Una pretensión quizá demasiado elevada ya que, si bien los chatbots son capaces de interactuar con una persona a partir de la entrada de texto, o incluso voz, la experiencia difícilmente se parecería a una conversación simple entre personas. De ahí que el desarrollo principal de este tipo de tecnología se haya dirigido a mejorar la experiencia conversacional: al uso de un lenguaje cada vez más parecido al humano que pueda ser percibido como más real y agradable.

En definitiva, los chatbots se plantean como una tecnología que confunda, que mienta y que elimine la separación entre personas y máquinas. Dilema que ya fue previsto por Turing al postular un estándar que diferenciase la «inteligencia» de un bot a la de una persona. En el hipotético caso de que no se pudiese diferenciar el comportamiento de una máquina con el de una persona, se habría alcanzado un nivel de imitación complejo con una serie de dilemas éticos nuevos e interesantes que hay que gestionar lo antes posible.

Está claro que la IA tiene importantes ventajas. En concreto, los chatbots proporcionan un servicio 24/7, rápido y para muchas personas, resolviendo las tareas más tediosas y automáticas eficazmente. En cambio, se habla poco de los inconvenientes que intentar reemplazar la atención personal pueda tener sin haber realizado un análisis de cuáles serían las funciones adecuadas, cómo gestionar errores, limitaciones, urgencias y otras situaciones problemáticas previamente o, por lo menos, intentando adelantarse a que pasen. Por encima de todo la IA no tiene moral, ni ética, ni deontología hasta que se programe y corrija constante y correctamente.

Cuando interactuamos con un chatbot estamos interactuando con una máquina programada para crear la ilusión de humanidad y eficiencia gracias a una lógica previamente establecida y que puede contener errores. Más aún si los procesos a los que responde el chatbot son cada vez más complejos como los sanitarios. Y es gracias al uso de cada vez más datos lo que crea la ilusión de una inteligencia real, por lo que el impacto sobre las personas es cada vez mayor. Por eso, el factor más importante y limitante de un chatbot no debería

ser si obtiene más o menos datos de calidad, sino si el algoritmo que analiza los datos arroja unas soluciones basadas en un modelo ético robusto, transparente y revisable. Recordemos que en breve las máquinas tendrán más datos que las propias personas que los aportan.

Existen principalmente dos tipos de chatbots. El primero corresponde a los dirigidos a tareas automáticas como la resolución de consultas de los usuarios, el diagnóstico a partir de cuestionarios, etc. Son interacciones muy estructuradas y cerradas, reduciendo la posibilidad de error y facilitando su uso, planificación y gestión, de ahí que sean los más utilizados. Así que, si existe una tecnología que imita sentimientos y conciencia, también debería imitar la moral y la ética. Sin embargo, no son este tipo de chatbots los que realmente provocan más preocupaciones sobre su uso.

El segundo tipo de chatbots son más complejos, ya que son capaces de interactuar con los usuarios y arrojar respuestas automáticas a las diferentes peticiones recibidas. Estos son los que pueden provocar discriminaciones que en según qué casos pueden ser realmente importantes, más si se utilizan, por ejemplo, para un cribado de salud. Estos chatbots predictivos tienen algoritmos desarrollados con IA para detectar conductas, pensamientos o estados de ánimo gracias a la cantidad ingente de datos que recogen (*big data* y *data mining*). Los datos no generan conocimiento por sí solos; es su correcto procesamiento el que lo consigue al crear determinados patrones mediante la comparación. Por eso, si se comparan con datos erróneos o sesgados, perpetuarán una discriminación injusta.

Imaginemos un *software* que implemente IA y que quiere detectar la probabilidad de padecer cierta enfermedad, pero que, para implementar el algoritmo, se ha basado en datos fiables pero recogidos únicamente de personas de cierta localidad, blancas, predominantemente hombres y con una edad superior a los 50 años, cuando la incidencia de dicha enfermedad es baja entre este colectivo. Claramente, solo reflejará las variables predictoras de dicha enfermedad cuando interactúen con personas de esas características. Es decir, la utilidad de dicho chatbot no es que sea negativa, es que directamente provoca una discriminación y puede empeorar la situación de algunas personas a las que el algoritmo informará que tienen un riesgo mínimo, siendo falso. En

otras palabras, una herramienta que podría ser muy útil para ahorrar tiempo, detectar fácilmente a personas en riesgo y prevenir la situación, tan solo sirve para una pequeña parte de la población que puede ser difícil de detectar en según qué proyectos o protocolos en los que se habrá basado el algoritmo inicialmente. Pero no olvidemos que el algoritmo se retroalimentará de los datos recogidos, sesgando aún más los resultados si no se corrige o retroalimenta con otras fuentes de información. Imaginemos también que dicho chatbot ha llegado a un nivel de desarrollo conversacional en el que no solo utiliza y reconoce algunas frases simples comprensibles, sino que tiene tal nivel de desarrollo que los usuarios son incapaces de diferenciarlos de un profesional de la salud; de hecho, creen que están interaccionando con uno o una, ya que en ningún lugar pone que no sea así. La mayoría de pacientes creará más firmemente en el diagnóstico erróneo de un algoritmo mal diseñado.

Otro ejemplo, no relacionado con la salud, podría ser el uso de un algoritmo para la vigilancia policial en un barrio considerado conflictivo. Dicho algoritmo recogerá más datos de ese barrio creando un bucle de retroalimentación pernicioso en el que cuanta más vigilancia policial más probabilidad de encontrar y registrar delitos, justificando los nuevos datos la necesidad de dicha vigilancia. Curiosamente los sesgos siguen dirigiéndose hacia colectivos vulnerables definiendo una realidad imperfecta y perpetuando una tecnología nociva, en mayor o menor medida, que analizará de forma incompetente por máquinas mientras que las personas privilegiadas seguirán siendo atendidas por personas.

Es decir, uno de los pilares más importantes del uso tecnológico en salud, que radica en acercar la atención de calidad y la eficacia al máximo de personas, si no se utiliza éticamente aumentará la actual brecha de acceso al sistema sanitario en lugar de reducirlo. Si las condiciones cambian, el modelo debería hacerlo también, siendo dinámico, ya que no es más que una representación simplificada y abstracta de ciertos procesos. Tenemos que valorar si, con la tecnología, en lugar de eliminar el sesgo humano lo que hemos hecho es camuflarlo, justificarlo y perpetuarlo.

Aun así, actualmente la IA no es capaz de desarrollar tareas múltiples complejas, pero sí una gran personalización y orientación a soluciones o tareas con-

cretas. Al ir paulatinamente superando esta limitación, cada vez será más frecuente encontrar que determinados procesos asistenciales integrados serán controlados y gestionados por algoritmos. Situaciones ética, deontológica y legalmente comprometidas, especialmente en relación con la cantidad, seguridad y calidad de los datos recogidos y la gestión de los posibles errores ocasionados y sus responsabilidades.

En cualquier caso, habría que enseñar ética al programador o programadora porque alguien tiene que ser responsable de los actos del chatbot. Es sumamente importante definir cuándo la persona deber tener el control final (*human-in-the-loop*, HITL), es decir, las personas deben seguir teniendo la última palabra sobre decisiones que afecten a la vida; cuándo se deben supervisar las acciones de una máquina (*human-on-the-loop*, HOTL), o cuándo la máquina podrá decidir autónomamente (*human-out-of-the-loop*, HOOTL). Y en muchos casos no sirve el ensayo y error, y hay que trabajar con una mentalidad predictiva lógica.

Conclusiones

La salud digital sin ética o deontología crea un problema donde debería solucionarlo. La tecnología puede ser la respuesta a muchos problemas, pero si es la propia tecnología la que crea el problema, la solución la encontraremos en una tecnología mejor, en una tecnología ética y que busque el bienestar global y no la ganancia económica de unos pocos.

Por suerte, los profesionales de la salud y el resto de profesionales involucrados en la salud digital están empezando a reaccionar a tiempo poniendo en el centro de la ecuación a la persona, no a la tecnología, pero también evaluando la pertinencia o no de las intervenciones y creando mecanismos de retroalimentación constante que detecte los sesgos, los posibles daños y perjuicios del uso de la tecnología mediante modelos dinámicos actualizados y no algoritmos cerrados.

En poco tiempo asistiremos al desarrollo de redes de comunicación 5G, lo que permitirá a la IA desarrollar funciones mejoradas y predicciones más rápidas, fiables y móviles. Seguramente, los asistentes personales serán la norma gene-

ral, con lo que muchas gestiones serán más eficientes y la sociedad estará más conectada, algo que sin un control ético estricto será un fracaso y retrasará su desarrollo aumentando las retenciones de los pacientes para usarlo. Por eso es tan importante iniciar, lo antes posible, una regulación completa de la implementación de la IA en la salud digital y conseguir programar algoritmos beneficiosos para el máximo de personas posible.

Bibliografía

- Acero-Rodríguez PD, Baena A. Ética y deontología de la telesalud mental. En: Martí Noguera JJ, coordinador. Manual de telesalud mental: Tecnologías digitales en la práctica clínica. Madrid: Pirámide; 2022. p. 29-41.
- Burzaco Samper M. Protección de datos personales y garantía de los derechos digitales, n.º del Estado, E. J. Ley Orgánica 3/2018, de 5 de diciembre (2020). Madrid: Dykinson; 2021. Disponible en: <https://doi.org/10.2307/j.ctv17hm980>
- Cebrián RP, de la Torre Martí M. Eficacia de la intervención por telesalud mental en adultos. En: Martí Noguera JJ, coordinador. Manual de telesalud mental: Tecnologías digitales en la práctica clínica. Madrid: Pirámide; 2022. p. 81-89.
- Garbowski M. A critical analysis of the Asilomar AI principles. *Zeszyty Naukowe. Organizacja i Zarządzanie/Politechnika Śląska*; 2018.
- Hsiung R. Suggested Principles of Professional Ethics for the Online Provision of Mental Health Services. *Studies in Health Technology and Informatics*. 2001; 84: 1296-1300. Disponible en: <https://doi.org/10.1089/153056201300093895>
- Mahoney N, Walaszek A, Caudill R. Incorporating Technology into the Psychiatric Residency Curriculum. *Academic Psychiatry*. 2018; 42(6): 847-851. Disponible en: <https://doi.org/10.1007/s40596-018-0978-3>
- Midkiff DM, Joseph Wyatt W. Ethical Issues in the Provision of Online Mental Health Services (Etherapy). *Journal of Technology in Human Services*. 2008; 26(2-4): 310-332. Disponible en: <https://doi.org/10.1080/15228830802096994>
- Real AP. La Declaración de derechos y principios digitales de la Comisión Europea. ¿Hacia un marco constitucional de los derechos digitales en Europa? *Derecho Digital e Innovación. Digital Law and Innovation Review*. 2021; (10): 2.

Robótica social: aspectos éticos y sociales

Miquel Domènech

Barcelona Science and Technology Studies Group (STS-b)

Introducción

Según nos advierten las estadísticas, la importancia de los robots en la economía y la sociedad ha ido aumentando progresiva y significativamente a lo largo de los últimos años. La mayoría de estos robots se han desarrollado para cubrir las necesidades de la industria. Se supone que los robots mejoran la calidad del trabajo al hacerse cargo de tareas peligrosas, tediosas o sucias que resultarían nada atractivas o poco aconsejables para los seres humanos. En este sentido, el robot industrial ha sido pensado como un sustituto del ser humano y su implantación en fábricas y talleres se ha llevado a cabo, normalmente, en jaulas o espacios protegidos a los que los humanos no deberían acceder. Esto ha comportado que en el contexto industrial no haya sido necesario tener en cuenta cómo gestionar la interacción entre humanos y robots; una vez colocado el robot, las personas han desaparecido.

Con el siglo XXI, sin embargo, estamos asistiendo a una importante aceleración en el desarrollo de la robótica, lo que ha llevado a los robots a salir de sus jaulas industriales y a colaborar con los humanos en el ejercicio de diferentes tareas. Ya no se trata de que operen en entornos altamente controlados como los laboratorios o las factorías, sino que se busca que se incorporen a la vida cotidiana de las personas. En este caso, se busca que lleven a cabo tareas en entornos poco o nada estructurados, con la posibilidad, además, de que se relacionen con humanos.

Atendiendo a estas consideraciones, se dice que estamos adentrándonos en una nueva era de la robótica –New Robotics, en inglés– que, a diferencia de la precedente, estaría centrada en los humanos –*human centered*– con robots diseñados para interactuar con humanos en entornos cotidianos.¹ Algunos países destacan especialmente en este cambio de lógica. El caso de Japón es bastante paradigmático y se señala el año 2005 como un momento clave en el viraje conceptual que se produce en el campo de la robótica y que posibilita que el énfasis de la producción de robots pase de la industria al sector servicios, lo que se expresa también en la implementación de políticas públicas que buscan el desarrollo de una sociedad donde robots y humanos coexistan armoniosamente.² Y, después, también tenemos el caso de Corea, que, ya hace prácticamente una década, preveía tener un robot en cada casa hacia 2020.³

En esta nueva era de la que hablamos los robots son pensados en un amplio abanico de entornos cotidianos, y el de los cuidados no ha sido una excepción. Es verdad que algunos estudios apuntan a que tanto pacientes como terapeutas pueden mostrar un cierto grado de miedo ante los robots.⁴ Ya hace unos años, Coeckelbergh *et al.*⁵ citaban el Eurobarómetro de 2012 para indicar que una gran parte de la población europea rechazaba la idea de usar robots para el cuidado de las personas. Concretamente, según los datos mencionados, el 60% de los ciudadanos de la Unión Europea decían entonces que los robots deberían estar prohibidos en el ámbito del cuidado de los niños, las personas mayores y las personas con discapacidad. Más recientemente, Gnams y Appel,⁶ en un estudio con datos que abarcaban hasta 2017, llegaban a la conclusión de que, de hecho, las actitudes de la población europea hacia los robots habían empeorado desde 2012.

Sin embargo, estas estadísticas no parecen haber desengañado a los investigadores y desarrolladores de tecnologías robóticas, que insisten en que el futuro que nos espera es uno en el que los robots se relacionan e interactúan con humanos. Esto supone, pues, la aparición de un nuevo abanico de preguntas e incertidumbres ante unas prácticas que son novedosas y de las que, por tanto, no tenemos precedentes donde ir a buscar respuestas. A continuación, me centraré en algunas de las preguntas e incertidumbres que se generan en un campo concreto: el de los cuidados. Para ello, primero presentaré los argumentos que se esgrimen para justificar la necesidad de robots para los cuidados, luego señalaré algunos usos que ya se están haciendo de robots sociales para el cuidado y, finalmente, plantearé algunas de las incertidumbres que generan, para acabar con algunas recomendaciones.

1. ¿Por qué robots para el cuidado?

Quizá la respuesta a la pregunta sobre el origen de la introducción de los robots en tareas de cuidado debe buscarse en una cuestión más general que tiene que ver con la relación entre el cuidado y la tecnología.

Bastante antes de que se pensara en robots para el cuidado, ya se había convertido en una cuestión de debate si la tecnología –en diferentes expresiones

no robóticas como los dispositivos de teleasistencia o la telemedicina– podría llegar a ser una solución para una esperada crisis del cuidado en un futuro no muy lejano.⁷ ¿De dónde surge la idea de una «crisis del cuidado»? Por qué la idea de que las sociedades occidentales están abocadas a una «crisis del cuidado» se ha convertido en un lugar común desde hace ya algunos años, no solo entre la literatura especializada, sino también en la prensa y los medios de comunicación generales.

Se citan estudios de diferentes agencias y organismos que insisten en datos que refuerzan esta idea de la crisis. Hay dos estadísticas que suelen sacarse a colación y que, de hecho, están interconectadas. Una tiene que ver con el progresivo envejecimiento de la población mundial. En este sentido, de acuerdo con el Departamento de Asuntos Económicos y Sociales de Naciones Unidas, «en 2018, por primera vez en la historia, las personas de 65 años o más en todo el mundo superaron en número a los niños menores de cinco años. Las proyecciones indican que para 2050 habrá más del doble de personas mayores de 65 años que niños menores de cinco años. Para 2050, el número de personas de 65 años o más en todo el mundo también superará el número de adolescentes y jóvenes de 15 a 24 años».⁸

La otra estadística hace referencia al aumento de las enfermedades neurodegenerativas, que va aparejado al envejecimiento de la población. Según datos publicados en septiembre de 2019 en su web, la OMS⁹ cifra la afectación de la demencia a nivel mundial en unos 50 millones de personas. Según esta misma fuente, cada año se registran cerca de 10 millones de nuevos casos y se calcula que entre un 5% y un 8% de la población general de 60 años o más sufre demencia en un momento determinado. La previsión, finalmente, es que el número total de personas con demencia llegue a los 82 millones en 2030 y a los 152 millones en 2050.

La conclusión a la que se llega a través de estadísticas como las que acabo de mencionar es que la humanidad envejece y cabría la posibilidad que en un futuro cercano no hubiera suficientes personas para llevar a cabo las tareas de cuidado de esta humanidad envejecida. La solución que se ofrece, entonces, es el desarrollo de dispositivos tecnológicos que palien este augurado déficit de personas cuidadoras.

Así pues, el desarrollo de robots de cuidado no es sino un paso más en un proceso de innovación tecnológica para encontrar soluciones a una posible falta de personas cuidadoras. En este contexto, el robot aparece como una solución más completa, más holística que, por ejemplo, un dispositivo de telealarma. La convergencia de la robótica con la inteligencia artificial permite pensar en la posibilidad de desarrollar robots sociales con la capacidad de analizar las necesidades de las personas usuarias a través de su interacción cotidiana con ellas y de actuar consecuentemente para dar respuesta.

Se trata, es cierto, de ideas que aún no están ni en pañales, que todavía necesitan bastante tiempo para que lleguen a concretarse en productos tecnológicos acabados y accesibles al gran público. Sin embargo, que detrás de ellas hay una firme voluntad de avanzar hacia soluciones viables para ciertos problemas de cuidado no se puede poner en duda. Y que los primeros pasos se están dando lo pone de manifiesto la presencia cada vez más frecuente en los medios de comunicación de informaciones respecto a distintas experiencias puestas en marcha, casi siempre como programas piloto.

Efectivamente, desde hace cierto tiempo en diferentes medios de comunicación se pueden encontrar referencias a usos pioneros de robots en contextos bien diferentes. Así, en un reportaje publicado en *ABC*, Díez explicaba que:

A los mayores de la residencia Shintomi, en Tokio, no les levanta cada mañana un despertador, sino Sota, un pequeño robot azul y blanco de brillantes colmillos negros, que preside sus mesitas de noche junto a las pastillas. Además de darles los buenos días y recordarles las horas de su medicación, este muñeco de treinta centímetros dispone de una cámara con rayos infrarrojos para detectar si los abuelos caen de la cama, lo que haría saltar una alarma en los móviles de los cuidadores. Desarrollado por Intellivoice, cuesta 300.000 yenes (2.300 euros) y es capaz de entablar una conversación sencilla con las personas mayores, para que no se sientan solas.

En esta misma línea, Connor Walsh, un experto en robótica de la Universidad de Harvard, en un reportaje de *QMayor Magazine*,¹⁰ tras subrayar el crecimiento imparable de las necesidades de cuidado, iba más lejos y acababa afirmando que «con el tiempo, todo el mundo será capaz de comprarse un robot, del mismo modo que ahora nos compramos un ordenador».

2. Robots de cuidado, ¿para qué?

¿Qué podrían aportar, pues, los robots sociales a las tareas de cuidado? Hay diferentes respuestas a esta pregunta, dependiendo, en gran parte, del optimismo de quien responda. Señalaré aquí solo algunos usos que ya se han experimentado, aunque sea de manera piloto, y que responden a ciertas necesidades conocidas que se cree que los robots podrían resolver.

2.1. Tratamiento del dolor y la ansiedad a través de la interacción de robots con niñas y niños

A nadie se le escapa que las estancias hospitalarias pueden llegar a ser una fuente de ansiedad muy importante para niños y niñas. Se trata de situaciones que normalmente son poco frecuentes, pero que son vividas con gran intensidad por los menores que, por otra parte, dada su inmadurez, suelen carecer de los recursos necesarios para una buena adaptación a los cambios que supone la hospitalización. El miedo al dolor, el ensañamiento del entorno familiar, la ruptura de las rutinas son factores que dificultan una buena actitud y un espíritu animado, ingredientes que, es bien sabido, contribuyen a la recuperación de las enfermedades.

Por eso, hoy en día, los hospitales pediátricos trabajan con afán para que las estancias hospitalarias infantiles sean lo más soportables posible, movilizándolo un ingente número de recursos con esta finalidad. No es extraño, pues, que la incorporación de robots para paliar la cuestión del dolor y la ansiedad haya sido contemplada, por parte de diferentes centros hospitalarios, como un nuevo recurso a considerar y que se haya respondido positivamente al interés de los grupos de investigación que trabajan en robótica social para niños. Así, diferentes experiencias piloto han sido llevadas a cabo utilizando robots humanoides¹¹ o mascotas robóticas¹² con resultados interesantes.

2.2. Atención a las necesidades escolares y comunicativas de niñas y niños hospitalizados

Otro de los aspectos que merece ser considerado al tratar la cuestión de las estancias hospitalarias de menores, especialmente cuando estas se prolongan

en el tiempo, es el seguimiento escolar y la comunicación con los padres. Como acabo de decir, los hospitales pediátricos no escatiman esfuerzos para facilitar la mejor de las estancias para sus pacientes. Sin embargo, el seguimiento escolar y la comunicación con personas fuera del hospital pueden representar un obstáculo mayúsculo.

Los niños y niñas pasan gran parte del tiempo en la escuela. Este entorno supone para ellos un lugar básico en su día a día. Allí aprenden, se educan y socializan. Ausentarse supone siempre un contratiempo. A través de robots que ocupan el lugar del niño o la niña hospitalizada, se han llevado a cabo algunas experiencias que permiten mantener su actividad escolar y continuar interactuando con sus compañeros de clase.¹³

Gracias a estos robots sustitutos, un niño o niña puede ir a clase, ver y escuchar a sus docentes habituales, interactuar con ellos y mantener una vida escolar lo más parecida posible a la que desarrolla en condiciones normales. Y no solo dentro del aula, sino también fuera de ella, ya que el robot también sirve para mantener contacto con sus iguales, a los que acompaña al patio o en otros momentos de recreo. El robot le permite, pues, no solo seguir con su aprendizaje, sino también mantener las relaciones sociales en el ámbito escolar.

2.3. Realización de tareas de apoyo

Hasta ahora me he referido a aplicaciones de robots sociales que conciernen especialmente a niños y niñas. Ahora bien, las personas adultas también son hospitalizadas y su cuidado a menudo requiere de la movilización de importantes recursos humanos para atender las diferentes necesidades. El uso de tecnología de soporte no es una novedad para la realización de ciertas tareas, como ayudarse de una grúa para levantar de la cama a personas con movilidad reducida.

Desde hace ya un tiempo, algunos desarrolladores de robots han señalado que este tipo de tareas más repetitivas y tediosas, como arreglar, alimentar o levantar pacientes podrían ser llevadas a cabo por robots.¹⁴ El desarrollo de brazos robóticos para realizar tareas de alimentación, por ejemplo, es bien conocido.

El robot BESTIC, por ejemplo, ha llegado a ser comercializado, aunque ha despertado algunas dudas sobre si se trata de una tecnología lo suficientemente madura para ser implantada.¹⁵

Se podría decir que una grúa o un brazo robóticos muy poco tienen de robot social. Ahora bien, el hecho de que, como he señalado, se trate de tareas que pueden parecer repetitivas o aburridas, no nos debe llevar a perder de vista que levantar a alguien de la cama o alimentar a una persona que no puede hacerlo por sí misma son tareas especialmente delicadas que según como se lleven a cabo pueden socavar la dignidad de la persona cuidada. En este sentido, la robótica asistencial está tendiendo a diseños más cercanos a la robótica social que permitan realizar estas tareas de manera menos mecánica.¹⁶ Por ejemplo, para el caso de la tarea de levantar y trasladar personas con movilidad reducida, se quiere abandonar diseños inspirados en grúas para proponer modelos que pretendan ser más amigables y con más prestaciones. Una muestra de ello sería el robot RIBA,¹⁷ que haría la operación de traslado de personas de manera análoga a como lo haría un ser humano.

2.4. Complemento para ciertas terapias

El uso de robots sociales para reforzar la labor de los terapeutas es un campo emergente en el que se están invirtiendo recursos considerables. Tal y como explican Heerink *et al.*,¹⁸ los avances en robótica social han permitido desarrollar robots más sofisticados, disponibles y asequibles, por lo que su uso en terapias para grupos con necesidades especiales, como las personas con demencia, los niños hospitalizados o los niños con autismo está siendo visto con especial interés por la comunidad médica.

El caso de niños y niñas con algún trastorno dentro del espectro autista (TEA) es contemplado con especial optimismo. La idea de que los robots pueden ser útiles en la comunicación con niños autistas por su capacidad para estimular su atención ha sido corroborada en diversos estudios.^{19,20} La dificultad que presentan los niños y las niñas con TEA para expresar emociones, por ejemplo, ha sido abordada con el uso de robots sociales que ayudan a los niños y niñas a aprender a reconocer expresiones faciales en sí mismos y en los demás. Por otro lado, el hecho de que a los niños y niñas con TEA les resulte especial-

mente estresante relacionarse con personas, esencialmente debido a la imprevisibilidad de la conducta humana, puede ser paliado eficazmente con el uso de robots sociales. Estos pueden ser diseñados para seguir patrones conductuales estables para que, en su interacción con los niños y las niñas, generen considerablemente menos ansiedad que los humanos.

3. Incertidumbres

En este apartado querría abordar algunas de las incertidumbres y temores que plantea el uso de robots sociales para el cuidado. Al tratarse de una innovación muy reciente, aún en vías de desarrollo, es evidente que las preguntas que surgen son muchas y variadas. Resultaría imposible, en el contexto de esta contribución, hacer un listado exhaustivo y tratar de darles una respuesta adecuada.

Es por ello que quiero centrarme únicamente en dos de las cuestiones que probablemente más recelos despiertan: ¿los robots pueden llegar a sustituir a los seres humanos en las prácticas de cuidado? ¿Es el cuidado que puede proporcionar un robot cualitativamente diferente del que dispensa un ser humano?

Para responder a la primera cuestión quiero plantear cómo entiendo que se llevan a cabo las prácticas de cuidado habitualmente. Fijémonos en el entorno hospitalario, por ejemplo. El que hallamos siempre es un entramado de cuidado donde hay una variedad de actores. En este entramado encontramos normalmente diferentes elementos, muchos de los cuales son, ciertamente, humanos (personal sanitario, personas voluntarias, familiares, quizá unos payasos...), pero también podemos detectar la presencia de diferentes recursos tecnológicos más o menos sofisticados (monitores multiparamétricos, camas eléctricas, termómetros...). Es todo el entramado en conjunto el que cuida. Sería una tarea complicada dilucidar quién tiene más agencia en este proceso de cuidado, ya que los diferentes componentes se necesitan los unos a los otros para producir el efecto de cuidado deseado. Es en este sentido que puede resultar tan conveniente una respuesta afirmativa como una negativa a la pregunta formulada. Cada vez que incorporamos un nuevo dispositivo a este

entramado –y lo mismo sería con cada nuevo actor humano– se reconfiguran las relaciones entre sus componentes, lo que puede hacer que algunos de ellos –o algunas de sus funciones– pasen a tener un papel residual o, incluso, dejen de tenerlo. En este sentido, una innovación cualquiera puede ciertamente reemplazar a un ser humano en un entramado de cuidado. Pero no tiene sentido pensar las innovaciones tecnológicas como meros dispositivos ideados para sustituir seres humanos. La incorporación de un aparato tecnológico debe considerarse a partir del efecto de cuidado que se quiere que el entramado produzca. Y esto puede suponer que, con la incorporación del dispositivo, se tenga que prescindir de un humano, ciertamente, pero también que se tenga que incorporar uno nuevo o, también, que haya que redefinir su papel. Así, por ejemplo, un robot social que entra en las habitaciones de un hospital interacciona con los pacientes y recoge datos sobre su estado, de ninguna manera convierte a la enfermera en un recurso prescindible, pero probablemente redefinirá sus funciones y rutinas. La pregunta que debemos plantearnos, por tanto, ante una innovación tecnológica sería más bien la siguiente: ¿cuidará mejor este entramado con la incorporación de este robot?

En cuanto a la segunda cuestión, su respuesta está, obviamente, relacionada con lo que acabo de decir respecto de la primera. Si entendemos que el cuidado es un efecto producido por todo un entramado heterogéneo de personas y tecnologías, plantear la existencia de un cuidado puramente humano hacia otro puramente robótico parece, en principio, un ejercicio reduccionista. Ahora bien, es cierto que la inquietud que pone de manifiesto la segunda pregunta se hace patente cuando, por ejemplo, se distingue entre la frialdad de la tecnología y sus productos y el calor del cuidado (que se presupone eminentemente humano). En este asunto, resultan especialmente interesantes las reflexiones de Pols y Moser,²¹ que cuestionan que haya una dicotomía entre la tecnología fría y el cuidado cálido humano. Para estas autoras, cálido y frío, al igual que racional y afectivo o tecnológico y sociable, no son opuestos, sino que son características que se alinean de manera diferente a diferentes prácticas de cuidado. La posibilidad de imaginar personas solitarias, privadas social y afectivamente por vivir entre tecnologías frías e instrumentales, insisten estas autoras, se sustenta en ignorar que la gente desarrolla relaciones afectivas con las tecnologías y que las tecnologías pueden ayudar a desarrollar

lazos sociales antes que cortarlos. De hecho, tal como lo plantean Pols y Moser, la incorporación de tecnología en entramados de cuidado puede servir también para fortalecer las relaciones entre pacientes y personal sanitario, de manera que los primeros se sientan no menos cuidados, sino más, aunque su interacción se pueda llevar a cabo esencialmente con un dispositivo tecnológico.

4. Recomendaciones

Finalmente, me gustaría cerrar esta contribución con algunas ideas a modo de recomendaciones que, desde las ciencias sociales, se pueden sugerir para poder llevar a cabo una toma de decisiones seria y fundamentada respecto al uso de robots sociales en el cuidado de las personas.

En primer lugar, creo que es importante evitar desatender miedos e incertidumbres como las que comentaba un poco más arriba. Fundamentadas o no, lo que inquietudes como estas evidencian es que el cuidado es una parte muy importante de nuestras vidas y que, por lo tanto, cualquier transformación debe ser abordada con prudencia y cautela. La manera en que cuidamos dice mucho del tipo de sociedad que somos, por lo que plantear cambios sustanciales en la manera de llevarlo a cabo no es un asunto menor.

Atender las reticencias de las personas significa abrir los procesos de toma de decisiones al conjunto de la ciudadanía y no dejarlos únicamente en las manos de las personas expertas. Difícilmente un dispositivo tecnológico servirá para cuidar mejor si aquellas personas para las que está concebido que ofrezca un servicio no entienden su razón de ser, lo ven como una amenaza o, simplemente, les parece inadecuado para esta tarea.

En segundo lugar, me parece conveniente extender la implicación ciudadana más allá del papel evaluador como usuarios destinatarios de las innovaciones. Demasiado a menudo, los innovadores acuden a nosotros, los científicos sociales, porque les ayudamos a que la gente acepte mejor sus propuestas tecnológicas. Esto hace patente que algunas innovaciones sean concebidas partiendo de lo que considero un error de base: innovar para la gente, pero sin la gente. La mejor manera de evitar el rechazo de las personas a las innovaciones tecnológicas es hacer que sean parte de ellas. Creo que es fundamental involucrar en

el desarrollo de robots sociales a los diferentes actores que, de una manera u otra, están implicados en su uso (pacientes, familiares, personas cuidadoras, profesionales de la salud...) y encontrar la manera de hacer posible la colaboración con ingenieros y diseñadores. En este sentido, querría señalar que no hay límites respecto al abanico de personas que pueden ser tenidas en cuenta.²²

En tercer lugar, no quería acabar sin antes insistir en un aspecto que ya ha sido sugerido, pero en el que quiero poner más énfasis: la tecnología no es neutra. Cada innovación tecnológica, y los robots sociales no son una excepción, incorpora valores, maneras de ver el mundo, promesas de sociedades futuras. Incorporar robots sociales a las prácticas de cuidado, no es bueno ni malo en sí mismo, pero tampoco es inocuo. Tendrá efectos en la forma de cuidar, en tanto que transformará relaciones, afectará a las identidades de las personas o cambiará los significados de determinadas situaciones, entre muchas otras posibilidades. Es básico, por tanto, que al plantearnos el uso de robots sociales para el cuidado nos formulemos preguntas sobre cuestiones como los valores que serán fortalecidos, las relaciones sociales que serán posibilitadas y las que serán dificultadas, o la manera en que mejorará el cuidado la incorporación de un robot social determinado. Y hacerlo siempre caso por caso, a partir de experiencias reales y/o plausibles. Los debates inspirados en fantasías futuristas pueden contribuir poco a la toma de decisiones serena y razonada.

Notas

1. Schaal S. The new robotics: towards human-centered machines. *HFSP Journal*. 2007; 1(2): 115-126.
2. Matsuzaki H, Lindemann G. The autonomy-safety-paradox of service robotics in Europe and Japan: a comparative analysis. *AI and Society*. 2016; 31(4): 501-517.
3. Borenstein J, Pearson Y. Robot caregivers: Harbingers of expanded freedom for all? *Ethics and Information Technology*. 2010; 12(3): 277-288.
4. Iosa M, Morone G, Cherubini A, et al. The three laws of neurorobotics: A review on what neurorehabilitation robots should do for patients and clinicians. *Journal of Medical and Biological Engineering*, 2016; 36(1): 1-11.
5. Coeckelbergh M, Pop C, Simut R, et al. A Survey of Expectations About the Role of Robots in Robot-Assisted Therapy for Children with ASD: Ethical Acceptability, Trust, Sociability, Appearance, and Attachment. *Science and Engineering Ethics*. 2016; 22(1): 47-65.
6. Gnambs T, Appel M. Are robots becoming unpopular? Changes in attitudes towards autonomous robotic systems in Europe. *Computers in Human Behavior*. 2019; 93: 53-61. Disponible en: <https://doi.org/10.1016/j.chb.2018.11.045>
7. Sánchez-Criado T, Jiménez M. ¿Personas mayores en autonomía conectada? Promesas y retos en la tecnologización del cuidado. *Revista Española de Investigaciones Sociológicas*. 2015; 152: 105-120.
8. United Nations D of E and SA. World Population Prospects 2019. Highlights. 2019. Disponible en: <http://www.ncbi.nlm.nih.gov/pub-med/12283219>
9. OMS. Demencia. [Notas descriptivas]. 2019. Disponible en: <https://www.who.int/es/news-room/fact-sheets/detail/dementia>
10. QMayor Magazine. Un nuevo sistema robótico puede ayudar a personas mayores y con discapacidad. 2016. Disponible en: <https://www.qmayor.com/tecnologia/personas-mayores-y-discapacidad/>
11. Jibb LA, Birnie KA, Nathan PC, et al. Using the MEDiPORT humanoid robot to reduce procedural pain and distress in children with cancer: A pilot randomized controlled trial. *Pediatric Blood and Cancer*. 2018; 65(9).
12. Díaz-Boladeras M, Angulo C, Domènech M, et al. Assessing pediatrics patients' psychological states from biomedical signals in a cloud of social robots. En: XIV Mediterranean Conference on Medical and Biological Engineering and Computing, MEDICON 2016, Paphos, Cyprus. IFMBE Proceedings (eds. E Kyriacou, S Christofides, and CS Pattichis); 2016. p. 1179-1184. Springer International Publishing.
13. Kristoffersson A, Coradeschi S, Loutfi A. A review of mobile robotic telepresence. *Advances in Human-Computer Interaction*. 2013.

14. Van der Plas A, Smits M, Wehrmann C. Beyond speculative robot ethics: A vision assessment study on the future of the robotic caretaker. *Accountability in Research*. 2010; 17(6): 299-315.
15. Nickelsen NCM. Imagining and tinkering with assistive robotics in care for the disabled. *Paladyn*. 2019; 10(1): 128-139.
16. Magnanato-Thalman N, Zhang Z. Assistive social robots for people with special needs. En: 2014 International Conference on Contemporary Computing and Informatics (IC3I). p. 1374-1380. IEEE.
17. Ding M, Ikeura R, Mori Y, Mukai T, Hosoe S. Measurement of human body stiffness for lifting-up motion generation using nursing-care assistant robot: RIBA. En: SENSORS. IEEE; 2013. p. 1-4.
18. Heerink M, Vanderborght B, Broekens J et al. New Friends: Social Robots in Therapy and Education. *International Journal of Social Robotics*. 2016; 8(4): 443-444.
19. Kozima H, Nakagawa C, Yasuda Y. Interactive robots for communication-care: a case-study in autism therapy. ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication. 2005: 341-346.
20. Albo-Canals J, Martelo AB, Relkin E, et al. A Pilot Study of the KIBO Robot in Children with Severe ASD. *International Journal of Social Robotics*. 2018; 10(3): 371-383.
21. Pols J, Moser I. Cold technologies versus warm care? On affective and social relations with and through care technologies. *ALTER-European Journal of Disability Research/Revue Européenne de Recherche sur le Handicap*. 2009; 3(2): 159-178.
22. Vallès-Peris N, Angulo C, Domènech M. Children's imaginaries of human-robot interaction in healthcare. *International Journal of Environmental Research and Public Health*. 2018; 15(5).

Relación de autores

- Antoni Baena, doctor en Psicología. Director de Programa del Máster Universitario en Salud Digital (MUSD). Estudios de Ciencias de la Salud. Universitat Oberta de Catalunya (UOC).
- Alicia de Manuel, investigadora del Observatorio de Ética en Inteligencia Artificial de Catalunya.
- Miquel Domènech, profesor de Psicología Social de la Universitat Autònoma de Barcelona. Fundador del Barcelona Science and Technology Studies Group (STS-b).
- Harry Farmer, investigador del Instituto Ada Lovelace de Londres.
- Felip Miralles, director de la Digital Health Unit. EURECAT.
- Juliana Ribera, Medical Consultant. Amalfi Analytics.

Títulos publicados

Cuadernos de Bioética

62. *Soledad no deseada en la era digital*
61. *Discapacidad y equiparación de derechos*
60. *Cuidarse en la sociedad entre pandemias*
59. *La atención a las necesidades sociales y sanitarias, ¿sumamos o dividimos?*
58. *Salud pública y COVID*
57. *Ciudades que cuidan, también al final de la vida*
56. *La salud pública, el género y la ética*
55. *Soledad, envejecimiento y final de la vida*
54. *Sexualidad y diversidad funcional*
53. *Nuevos escenarios en salud e investigación clínica*
52. *Ética y promoción de la salud. Libertad-paternalismo*
51. *Una mirada ética en la gestión de conflictos*
50. *Pensar la maternidad*
49. *Publicidad y salud*
48. *Prioridades y políticas sanitarias*
47. *Ética y donación de plasma: una mirada global*
46. *Comités de Ética y consultores clínicos: ¿complemento o alternativa en la ética asistencial?*
45. *CRISPR... ¿debemos poner límites a la edición genética?*
44. *Crisis y salud mental en niños y jóvenes: ¿causa o consecuencia?*
43. *¿Debemos revisar el concepto de muerte?*
42. *Iatrogenia y medicina defensiva*
41. *Eutanasia y suicidio asistido*
40. *Ethical aspects of research with children*
39. *Discapacidad, nuevos enfoques y retos éticos a la luz de la Convención de la ONU*
38. *Ética, salud y dispendio del conocimiento*
37. *Determinantes personales y colectivos de los problemas de la salud*
36. *Ética y altruismo*
35. *Treinta años de técnicas de reproducción asistida*
34. *Ética de la comunicación corporativa e institucional en el sector de la salud*
33. *Alcance y límites de la solidaridad en tiempos de crisis*
32. *Ética y salud pública en tiempos de crisis*
31. *Transparencia en el sistema sanitario público*
30. *La ética del cuidado*
29. *Casos prácticos de ética y salud pública*
28. *La ética en las instituciones sanitarias: entre la lógica asistencial y la lógica gerencial*
27. *Ética y salud pública*
26. *Las tres edades de la medicina y la relación médico-paciente*
25. *La ética, esencia de la comunicación científica y médica*
24. *Maleficencia en los programas de prevención*
23. *Ética e investigación clínica*
22. *Consentimiento por representación*
21. *La ética en los servicios de atención a las personas con discapacidad intelectual severa*
20. *Retos éticos de la e-salud*
19. *La persona como sujeto de la medicina*
18. *Listas de espera: ¿lo podemos hacer mejor?*
17. *El bien individual y el bien común en bioética*
16. *Autonomía y dependencia en la vejez*
15. *Consentimiento informado y diversidad cultural*

14. *Aproximación al problema de la competencia del enfermo*
13. *La información sanitaria y la participación activa de los usuarios*
12. *La gestión del cuidado en enfermería*
11. *Los fines de la medicina*
10. *Corresponsabilidad empresarial en el desarrollo sostenible*
9. *Ética y sedación al final de la vida*
8. *Uso racional de los medicamentos. Aspectos éticos*
7. *La gestión de los errores médicos*
6. *Ética de la comunicación médica*
5. *Problemas prácticos del consentimiento informado*
4. *Medicina predictiva y discriminación*
3. *Industria farmacéutica y progreso médico*
2. *Estándares éticos y científicos en la investigación*
1. *Libertad y salud*

Informes de la Fundació

6. *La interacción público-privado en sanidad*
5. *Ética y biología sintética: cuatro corrientes, tres informes*
4. *Las prestaciones privadas en las organizaciones sanitarias públicas*
3. *Clonación terapéutica: perspectivas científicas, legales y éticas*
2. *Un marco de referencia ético entre empresa y centro de investigación*
1. *Percepción social de la biotecnología*

Interrogantes éticos

5. *Pedagogía de la Bioética. Comunicaciones*
4. *Repensar el cuerpo*
3. *La subrogación uterina: análisis de la situación actual*
2. *Afectividad y sexualidad. ¿Son educables?*
1. *¿Qué hacer con los agresores sexuales reincidentes?*

Para más información: www.fundaciongrifols.org

FUNDACIÓ
VÍCTOR
GRÍFOLS
i LUCAS