INTELIGENCIA ARTIFICIAL ÉTICA EN SANIDAD

RECOMENDACIONES PARA LA ADOPCIÓN DE IA RESPETUOSA, TRANSPARENTE, SEGURA Y JUSTA

digitales_

BIENVENIDA

Una ayuda extraordinaria en el manejo de la incertidumbre

La Medicina es la ciencia de la incertidumbre y el arte de la probabilidad. Esta frase extraída del discurso de William Osler a principios del Siglo XX a sus alumnos de medicina, para muchos, entre los que me encuentro, es la definición más ajustada a la realidad y que sigue siendo válida en nuestros días.

Sin ese foco en la probabilidad, en los datos, no se habrían producido en la Historia muchos de los descubrimientos y avances en el ámbito de las ciencias de la salud. Avances que han repercutido de manera determinante en el objetivo final de estas ciencias: Prevenir y Curar.

No obstante, la realidad nos muestra lo lejanos que estamos aún de asegurar tanto la prevención como la curación, y el ejemplo que estamos viviendo con la pandemia del COVID-19 es suficientemente demostrativo.

En el documento que se presenta, intentamos aportar respuestas a una de las grandes incógnitas: cómo respetar la propiedad de los datos clínicos y, al mismo tiempo, maximizar las oportunidades de su utilización. De un lado, sin esos datos no es posible la investigación en ciencias de la salud. De otro, resulta esencial garantizar los derechos fundamentales de la privacidad y la confidencialidad.

Hoy día, una parte importante de las decisiones clínicas las basamos en los algoritmos deterministas o infalibles, los cuales están presentes en el software de la mayoría de las herramientas y aparatos con los que exploramos o intervenimos a los pacientes. Estos algoritmos son posibles y lo seguirán siendo en la medida que los investigadores dispongan de mayor cantidad de datos y del incremento de su calidad.

La nueva dimensión de los algoritmos -los conocidos como predictivos- necesitan aún más número de datos para que la probabilidad se acerque cada vez más a la certeza.

Los que trabajamos en el ámbito de la salud sabemos que, en la medida que estas herramientas avancen, la medicina será más ciencia que arte y nos podremos adelantar a la probabilidad y al azar.

Dr. Eduardo Vigil Martin, Chief Medical Officer Health NTT Data Company

Este trabajo ha sido elaborado por la Línea de Trabajo de Transformación Digital de DigitalES -Asociación Española para la Digitalización-.
Agradecimientos:
Huawei, IBM, Nae, NTT DATA España, PwC Tax & Legal, Sopra Steria y Telefónica.
Imágenes: Unsplash.

“Un informe que asienta las bases para el desarrollo de una Inteligencia Artificial ética y segura en el ámbito sanitario”

El impacto de la tecnología en las sociedades modernas está dejando una huella clara en los diferentes agentes sociales e, ineludiblemente, en las instituciones públicas y las compañías privadas. En este contexto, la Inteligencia Artificial es posiblemente la tecnología con mayor potencial para reformular cómo desarrollamos las actividades humanas. Ya no se trata de una nueva transformación digital, sino que estamos en un proceso de redefinición de los modelos comerciales y sociales.

Estamos, por tanto, en un momento de grandes oportunidades, pero también de grandes riesgos si no se pone a las personas y su protección en el centro del nuevo modelo tecnológico. Con el objetivo de reducir y prevenir el posible impacto negativo de la IA, la Comisión Europea publicó el pasado 21 de abril la propuesta de marco regulatorio, conocido como el “AI Act”, documento que plantea un *“instrumento legislativo horizontal que aplique un enfoque proporcional basado en el riesgo más un código de conducta para los sistemas de IA que no sean de alto riesgo".*

Esta propuesta de regulación liga el concepto de riesgo a potenciales vulnerabilidades a la salud y seguridad de las personas, elevando así la salud como uno de los principales elementos a proteger. En el ámbito sanitario cobra por tanto especial importancia el hecho de desarrollar soluciones tecnológicas éticas y responsables.

En NTT DATA denominamos a este desarrollo “de Platón a Python”. Desde esta aproximación, ayudamos a las organizaciones a definir una estrategia responsable para llevar los principios éticos a un entorno de desarrollo tecnológico, respondiendo a preguntas clave tales como: ¿Están definidos los mecanismos para asegurar la supervisión humana sobre la discrecionalidad de la IA?, ¿existe una delimitación clara sobre quién (humano o máquina) es responsable sobre las decisiones tomadas por la IA? ¿Es posible revertir una decisión que ha tomado un sistema de IA?

Este tipo de cuestiones deben plantearse desde la ideación de las iniciativas y han ser aplicadas de manera transversal a lo largo de todo el ciclo de vida del dato y del algoritmo.

Es por ello por lo que desde NTT DATA, en colaboración con DigitalES y con otras compañías líderes en el mercado español y global, hemos desarrollado este informe con el fin de aportar guías y propuestas para desarrollar una Inteligencia Artificial ética y segura en el ámbito sanitario. Gracias a la colaboración de numerosos expertos y profesionales, a partir de su experiencia, hemos hecho un recorrido por los principios éticos de la IA, explicando sus desafíos y realizando más de una veintena de propuestas para su cumplimiento.

David Pereira, Socio Responsable de IA en NTT DATA


digitales_

INTRODUCCIÓN

Cada año, la esperanza de vida aumenta en nuestro país aumenta en torno a 0,5 años. Siguiendo esta tendencia, antes de 2030 habremos alcanzado una esperanza de vida media de 100 años. Y lo que es más ilusionante todavía: lo haremos, manteniendo una elevada calidad de vida.

En España, al igual que en la mayoría de los países del llamado Primer Mundo, la esperanza de vida se encuentra estrechamente vinculada a la calidad asistencial, y ésta, a su vez, al avance científico y tecnológico. La aparición del virus SARS-CoV-2 que provoca la enfermedad COVID-19 ha tenido un impacto social y económico dramático, pero al mismo tiempo ha evidenciado la necesidad -geoestratégica, incluso- de apalancar el progreso en la ciencia, la innovación y la tecnología.

Durante la pandemia del COVID-19, forzados por las sucesivas políticas de confinamiento y restricciones de aforo, la transformación digital del sector sanitario en España se ha acelerado sensiblemente. La barrera cultural de la telemedicina prácticamente ha desaparecido, y los propios profesionales sanitarios demandan la implantación urgente de mejoras tecnológicas que reviertan en eficiencias y destensionen un sistema al límite de su capacidad.

En este ya de por sí difícil contexto, también se ha puesto a prueba la solidez y robustez de los sistemas tecnológicos ya implantados. Concretamente, de los sistemas de inteligencia artificial (IA). **La IA en el ámbito sanitario presenta una serie de particularidades éticas relacionadas con 1) unas necesidades de privacidad de datos y de robustez de los algoritmos superiores a otros ámbitos; 2) una necesidad muy clara de establecer la responsabilidad (accountability)**, sin que ello disuada a los profesionales médicos de emplear estas herramientas.

Para ello, resulta esencial realizar un análisis racional, basado en la evidencia. En la actualidad existen numerosos casos y ejemplos de cómo la Inteligencia Artificial aporta beneficios a la sociedad e impulsa la innovación de nuestros sectores. **El estudio detenido de tales ejemplos es el mejor modo en el que los legisladores pueden decidir sobre pasos futuros.** Creemos que la magnitud y relevancia de esta tecnología implementada en un campo tan sensible como el sanitario, se tiene que hacer atendiendo a las recomendaciones de los profesionales y con una temporalidad acorde a la importancia e impacto que conlleva.

Desde **DigitalES, Asociación Española para la Digitalización**, queremos contribuir a acelerar la incorporación de tecnologías y modelos de IA en el ámbito sanitario. Una incorporación que sea operativa y efectiva, en tanto revierta verdaderamente en una mejora de la calidad asistencial.

La experiencia acumulada por las empresas que forman parte de esta patronal, así como de los colaboradores y expertos para la elaboración del presente

documento, nos permite extraer una serie de aprendizajes útiles para Administraciones competentes, centros sanitarios y profesionales, tanto del ámbito médico como del tecnológico.

Por todo lo anterior, el lector encontrará al final de cada capítulo una breve enumeración de propuestas para el buen uso de la IA, y de la tecnología en sentido amplio, en el ámbito sanitario. Recomendaciones de medidas y actuaciones, en definitiva, para el desarrollo de una **IA respetuosa, transparente, segura y justa**.

¿QUÉ ES LA IA?

El concepto de Inteligencia Artificial ha ganado presencia en nuestro día a día, aumentando paulatinamente su relevancia tanto en nuestras conversaciones e interacciones como en su aplicabilidad y utilidad en el mundo de los negocios. Su concepto, aunque en apariencia obvio, genera diferencias y matices dependiendo del experto o emisor. La Unión Europea define la Inteligencia Artificial como:

"Habilidad de una máquina de presentar las mismas capacidades que los seres humanos, como el razonamiento, el aprendizaje, la creatividad y la capacidad de planear"

pero la evolución tecnológica es rápida y ofrece ya importantes capacidades y utilidades en diferentes sectores de actividad.

En la actualidad, la IA permite que los sistemas tecnológicos perciban su entorno, se relacionen con él, resuelvan problemas y actúen con un fin específico. La máquina recibe datos (ya preparados o recopilados a través de sus propios sensores, por ejemplo, una cámara), los procesa y responde a ellos. Los sistemas de IA son capaces de adaptar su comportamiento en cierta medida, analizar los efectos de acciones previas y de trabajar de manera autónoma.

Las aplicaciones que ofrece la IA tienen un potencial inimaginable en resolución y comprensión de problemas; se plantea como una tecnología sin límites. Uno de los sectores más destacados en el uso y aplicación de la IA es el sanitario. La IA permite monitorizar volúmenes ingentes de datos, buscar patrones y tendencias para encontrar soluciones que atiendan a la salud de las personas, mejorando la precisión y rapidez en aplicaciones tales como diagnósticos médicos, apoyo a tratamientos de pacientes o incluso ampliar las capacidades de telemedicina.

La pandemia del COVID-19 ha aumentado las expectativas para la IA en el sector sanitario, no sólo por su potencial para abordar futuros eventos similares, sino también para aportar a la investigación de soluciones de salud basadas en algoritmos, que permitan acelerar las fases de ensayo clínico.

En el sector de la sanidad, la aplicación de la IA cubre un amplio espectro de soluciones, desde la interpretación de imágenes de diagnóstico médico, lo cual

agilizaría el sistema y supondría una eliminación significativa de la probabilidad de error, hasta el seguimiento de enfermos críticos para paliar eventos inesperados o adversos. En todos los casos, la aplicación de la IA resulta en un incremento de la eficiencia de los resultados.

Otro de los valores en alza de la IA reside en su potencial para abordar futuras pandemias. La IA se alimenta de datos, y cuando dentro de años tengamos bases de datos que abarquen una temporalidad amplia de información sobre la crisis del coronavirus, la IA será útil para realizar diagnósticos tempranos y hacer predicciones de brotes, no sólo del COVID-19, sino de futuras enfermedades.

La IA permite extraer patrones, detectar anomalías o proyectar escenarios futuros a partir del análisis de grandes cantidades de datos. Cuando la muestra de datos es reducida, es evidente que los resultados podrían no ser los deseados, al menos a corto plazo. No en vano, toda la llamada 'ciencia de datos' obtiene su mejor potencialidad en entornos de rigor estadístico. Por eso, si bien tendemos a hablar de IA como una tecnología única, en realidad cada modelo de IA es único y ha sido diseñado para resolver un problema específico.

La idea de una "IA global", hoy por hoy, es más cercana al cine que a la realidad.

INTRODUCCIÓN A LA DIMENSIÓN ÉTICA DE LA IA

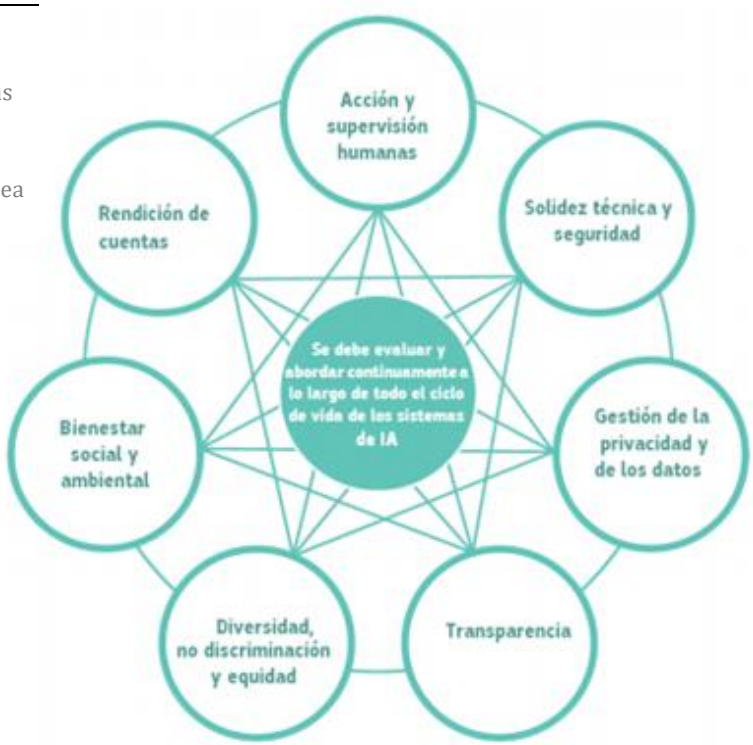
La aceleración en la implantación de la Inteligencia Artificial no solo implica beneficios, sino que acarrea una serie de riesgos que deben ser analizados en profundidad debido a las áreas tan sensibles que perpetra. Por ejemplo, la gestión y manejo de datos puede ser un tema susceptible de conflicto con el derecho de las personas, planteando cuestiones tanto legales como éticas. También, la posibilidad de tomas de decisiones sesgadas por el sistema de Inteligencia Artificial o la opacidad o falta de comprensión en procesos que llevan a las tomas de decisiones. Una de las preguntas claves es: cuando un sistema toma una decisión en base a datos y esto origina un resultado en salud, ¿quién es responsable de sus efectos?

El sector sanitario posee una criticidad que debe ser valorada a la hora de implantar cualquier solución de Inteligencia Artificial. Es decir, la dimensión ética tiene que ser parte integral de su desarrollo, pero también de su introducción y posterior seguimiento.

Ante esta nueva realidad, la Unión Europea ha enfrentado este reto con una batería de propuestas e iniciativas regulatorias que pretenden fomentar un avance tecnológico, al tiempo que se garantice la seguridad y proteja los derechos humanos. Una de las primeras iniciativas puesta en marcha por la Comisión Europea en el año 2019, fue la publicación por parte del Grupo de Expertos de Alto nivel de las *Directrices*

Gráfico 1.

Directrices éticas para una IA de confianza. Comisión Europea



Éticas para una IA de confianza[1]. Previamente, en diciembre de 2018, la Comisión ya había publicado un borrador para consulta pública, que recibió más de 500 respuestas. La versión final sufrió varios cambios, gracias a las aportaciones por la sociedad civil, instituciones y empresas, entre las cuales se encuentran algunas de las empresas asociadas a DigitalES. Estas Directrices sentaron las bases sobre las cuales se ha desarrollado las propuestas de regulación siguientes, habiendo dejado claro la intención de que dichas recomendaciones sean aplicables más allá de Europa y fomenten un debate global sobre esta cuestión.

Estas Directrices se dividen en tres partes: **los fundamentos de una IA fiable** expresados a través de cuatro principios éticos, **la realización de una IA fiable** mediante siete requisitos clave, y el **análisis de una IA fiable** que conduce a una lista de evaluación con preguntas específicas.

En concreto, se mencionan siete requisitos clave para el desarrollo de una IA fiable: (1) acción y supervisión humanas, (2) solidez y seguridad técnica, (3) privacidad y gobernanza de datos, (4) transparencia, (5) diversidad, no discriminación y equidad, (6) bienestar ambiental y social y (7) responsabilidad.

Posteriormente, en el año 2020, se presentó el "Libro Blanco sobre la inteligencia artificial: enfoque europeo orientado a la excelencia y la confianza"[2], donde se señala que para aprovechar las oportunidades que ofrece la inteligencia artificial y abordar los retos que presenta,

[1] https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[2] https://eur-lex.europa.eu/legal-content/ES/TXT/?qid=1603192201335&uri=CELEX%3A52020DC0065

la UE debe actuar conjuntamente y determinar de qué manera, a partir de los *valores europeos*, promoverá su desarrollo y adopción.

Actualmente, la Comisión ha presentado una propuesta de *Reglamento por el que se establecen normas armonizadas sobre inteligencia artificial*[3] donde la CE ha decidido no centrarse en la regulación de la tecnología sino en los "usos" de esa tecnología, estableciendo una clasificación o graduación de los riesgos que pueden causar: riesgo inaceptable, alto riesgo, riesgo limitado y riesgo mínimo. **Una aproximación adecuada, desde el punto de vista de DigitalES, dada la posibilidad de desactualización inmediata.**

Ante este impulso normativo, en proceso de desarrollo, las empresas e instituciones líderes han demostrado un alineamiento temprano atendiendo a la importancia de generar una IA de confianza. Los principios éticos para la implantación, desarrollo y uso de la IA se muestran como los cimientos sobre los que edificar unos sistemas sólidos que permitan el crecimiento y aprendizaje constante. Por las peculiaridades del sector sanitario, algunos de esos principios se erigen como el eje sobre el que pivotar los proyectos.

Es el caso de la llamada **"Explicabilidad de los datos"**, que permite a los profesionales sanitarios entender el modo en que los sistemas de IA alcanzan sus conclusiones y es por tanto esencial para la confianza de los médicos en estos sistemas. O la prevención de daño, ya que la IA se debe siempre utilizar en entornos seguros y controlados bajo la supervisión humana. En el sector sanitario, resulta crucial asegurar que una tecnología no genere daño alguno a los seres humanos en lo referente a su salud o a su privacidad, teniendo en cuenta el carácter particularmente sensible de la información personal sobre salud.

También resulta de gran trascendencia la **"Robustez de los sistemas"**, ya que cualquier error podría impactar y afectar gravemente a la vida de las personas. Además, la representatividad de los datos, mitigando sesgos potencialmente discriminatorios que conlleven a resultados menos favorables para determinados grupos y colectivos vulnerables.

MISIÓN DE PAÍS: LIDERAR EL DESARROLLO DE LA IA ÉTICA

España tiene una gran oportunidad de erigirse en una referencia internacional de IA ética en el ámbito sanitario. Contará para ello con dos grandes herramientas, la **Estrategia Nacional de Inteligencia Artificial (ENIA),** que contempla una inversión pública de 600 millones de euros en el periodo 2021-2023, y el **PERTE para una Sanidad de Vanguardia**, que movilizará una inversión público-privada de casi 1.500 millones de euros. Entre otros proyectos, la ENIA prevé la creación de **espacios de datos sectoriales** para industrias como la sanitaria. De su lado, el citado PERTE espera facilitar la transferencia del I+D+I académico al sector industrial y el impulso de la IA.

[3] https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=COM:2021:206:FIN


digitales_

CASOS DE USO

A continuación, enumeramos los principales casos de uso de la IA en el ámbito sanitario de países comparables con España, en términos de recursos y de alcance poblacional del propio sistema.

1.- TRATAMIENTOS Y CUIDADOS PERSONALIZADOS

Se están empezando a desarrollar las primeras soluciones que permiten encontrar el mejor plan de tratamiento personalizado para cada paciente, reduciendo los costes y aumentando la efectividad de dichos tratamientos.

Un ejemplo lo encontramos en la compañía americana **GNS Healthcare**, que desarrolla modelos de *machine learning* para determinar los tratamientos más efectivos a las características de cada paciente. Para lograrlo, trabajan en colaboración con las principales aseguradoras de salud y compañías biofarmacéuticas, utilizando sus datos anonimizados para simular la respuesta de los pacientes a distintos tratamientos. En España, **Sanitas** ha llevado a cabo el seguimiento en remoto de pacientes COVID a través de un sistema de IA. En caso de parámetros anómalos, el médico podía contactar con ellos a través de videoconferencia, reduciendo así considerablemente los reingresos.

Watson, el sistema de inteligencia artificial desarrollado por **IBM**, ha hecho que su capacidad para aprender lo convierta en un superordenador con aplicaciones en campos muy diversos. Entre ellos, la medicina es una de las áreas que mejor está aprovechando sus habilidades. En oncología, los especialistas del Memorial Sloan-Kettering de EEUU lo han entrenado para que pueda analizar historiales médicos y ayudar al médico a definir tratamientos más personalizados basados en estudios científicos.

En este tipo de aplicaciones, resulta de especial trascendencia garantizar el requisito de **"Privacidad de los datos"** por la tipología de datos personales que se deben manejar.

2.- ASISTENCIA AL DIAGNÓSTICO MÉDICO

En la **Universidad de Valencia** están entrenando a máquinas para ayudar a los médicos a realizar mejores diagnósticos. En concreto, trabajan en un sistema de inteligencia artificial capaz de detectar el cáncer de pecho en mamografías usando básicamente la misma tecnología que los físicos de partículas han utilizado para detectar el bosón de Higgs en el CERN de Ginebra.

Cada vez son más los ejemplos que encontramos de sistemas que utilizan el análisis de imágenes mediante IA para realizar diagnósticos médicos. Por un lado, pueden resultar muy útiles cuando se trata de reconocer patologías muy comunes, como un cáncer de pecho, un cáncer de piel o una retinopatía diabética,

para los que existe una importante base de datos históricos de los que aprender. Pero, por otro lado, resultan especialmente útiles a la hora de realizar diagnósticos tempranos de determinadas patologías que todavía están en una fase muy incipiente y que pueden ser difíciles de detectar por un profesional médico.

Existen en la actualidad multitud de soluciones para estos fines, como **Ezra**, que se centra en la detección temprana de cáncer a partir del análisis de la imagen de una resonancia magnética de cuerpo entero, o como **SkinVision**, que permite una detección temprana del cáncer de piel a partir de fotos tomadas por el propio paciente con su teléfono móvil.

Ya hay empresas, como la estadounidense **Heartflow**, a las que se les puede enviar un escáner realizado en cualquier lugar del mundo para obtener un diagnóstico remoto utilizando *deep learning.*

En el caso de la retinopatía diabética, el cribado de imágenes del ojo de personas diabéticas tiene una fiabilidad estimada superior al 95%, según diferentes estudios.

De nuevo, en este tipo de soluciones es importante garantizar el principio de "**Acción y supervisión humana**" para que el diagnóstico definitivo recaiga en un profesional médico, pero también cobran importancia otros requisitos como el de "**Solidez técnica y seguridad**" dado que buscaremos que nos den respuestas fiables y reproducibles, esto es, la misma respuesta ante la misma situación.

3.- MEDICINA PREVENTIVA Y AUTOCUIDADO

La IA puede usarse asimismo de forma preventiva, para predecir enfermedades o eventos de salud antes de que ocurran, y promocionando hábitos saludables acordes a cada persona. Por ejemplo, una tecnología de IA podría establecer el riesgo relativo de padecer ciertas enfermedades como la diabetes en función de factores determinantes como los antecedentes familiares o el estilo de vida.

Uno de los grandes dilemas o desafíos de la IA en el ámbito diagnóstico es su escalabilidad, es decir, si la IA entrenada para su uso en un contexto determinado se puede utilizar de forma precisa y segura en una región geográfica o contexto diferentes.

La IA podría eventualmente cambiar la forma en que los pacientes autogestionan

El sector sanitario en España ha avanzado mucho en la **interoperabilidad** de sistemas y en el manejo de herramientas digitales por parte de los profesionales sanitarios. Durante los últimos dos años, hemos visto una expansión rápida de la telemedicina y de los diagnósticos por imagen a distancia, repercutiendo en la eficiencia del sistema y en la calidad asistencial. **La IA permitirá transicionar hacia una medicina personalizada y preventiva**; un sistema sanitario más sostenible, que se ajuste al aumento de la esperanza de vida y al creciente peso de las afecciones crónicas sobre las agudas, y que al mismo tiempo que prevenga de la asunción de conclusiones equivocadas en las que también puede incurrir un sistema computacional.

sus propias afecciones médicas, particularmente aquellas de carácter crónico. Combinada con otras tecnologías, como el Internet de las Cosas y los dispositivos *wearables*, pueden ayudar a desarrollar servicios de telemedicina más 'inteligentes'. Los *wearables* resultan especialmente útiles para el seguimiento de los mayores.

Existen ya diferentes soluciones digitales que, bien mediante un cuestionario o a través de un *chatbot* inteligente en los casos más punteros, el usuario puede introducir una serie de datos describiendo su estado de salud, a partir de los cuales un modelo determinará la patología más probable y la especialidad a la que se recomienda acudir, pudiendo incluso contactar por chat directamente con el especialista y/o programar una videoconsulta con el médico. Un ejemplo de estas tecnologías lo encontramos en la plataforma digital de servicios de salud SAVIA de **Mapfre**.

Entre sus principales ventajas destacaríamos la facilidad de uso y su disponibilidad desde cualquier dispositivo electrónico (móvil, tablet, PC…), así como la rapidez para poder identificar una patología grave que requiera de una rápida intervención médica.

Sin embargo, es importante destacar que este tipo de soluciones todavía no proporcionan un diagnóstico médico como tal, en línea con el requisito de **"Acción y supervisión humana"** para una IA fiable, dado que su nivel de precisión todavía requiere de la validación de un profesional médico.

Los *chatbots* se han demostrado eficaces, asimismo, en el ámbito de la salud mental. **Sanitas**, por ejemplo, ha desarrollado un Asistente Virtual Psicológico, permite a cualquier persona acceder a una evaluación psicológica y a distintos recursos de salud a través de cualquier dispositivo con conexión a internet utilizando su voz o la escritura. La herramienta, validada validado por los psicólogos y psiquiatras de los hospitales de Sanitas, consigue superar la barrera que representa el estigma social. En esta misma línea, la compañía ha creado *chatbots* enfocados a ofrecer orientación en materia de reproducción asistida, un test de alzheimer, información sobre tratamiento de obesidad, orientación sobre salud íntima, etc.

4.- INVESTIGACIÓN BIOMÉDICA Y DESARROLLO DE NUEVOS FÁRMACOS

La tecnología hace posible el progreso de las técnicas clínicas y de la investigación. Durante la pandemia, este ámbito ha dado un salto muy importante, siendo las vacunas frente al virus el mejor exponente de ello.

Las compañías farmacéuticas, cada vez con más frecuencia, están aprovechando la IA para el diseño de nuevos fármacos. Éste es el caso, por ejemplo, de **NuMedii** y su solución AIDD (Artificial Intelligence for Drug Discovery), que aplica la IA para descubrir correlaciones y efectos entre los compuestos químicos de los medicamentos y diferentes enfermedades.

En este tipo de soluciones, la mayor complejidad reside en el manejo de los enormes volúmenes de datos y la disparidad de éstos que se deben analizar, si bien los avances en Machine Learning y, especialmente, en Deep Learning ya están

dando sus frutos, consiguiendo reducir los costes y el tiempo necesario de investigación para nuevos fármacos, en particular para enfermedades raras y los tipos de cáncer menos frecuentes.

Por la particularidad de estas soluciones y el resultado que buscan, que es el de nuevos fármacos seguros para la población, es de especial importancia garantizar el principio de "**Rendición de cuentas**" (*accountability*) por la necesidad de poder auditar los resultados y minimizar los posibles efectos negativos de estos fármacos.

Otra de las aplicaciones de la IA para la investigación en salud se encuentra en el campo de la genómica. Dado que el ser humano cuenta con aproximadamente tres mil millones de pares de bases de ADN, la medicina genómica es una disciplina emergente que puede beneficiarse enormemente de estas tecnologías.

5.- GESTIÓN OPTIMIZADA DE LOS HOSPITALES Y MEJORA DE LA CALIDAD ASISTENCIAL

También es aplicable la IA, por supuesto, a la hora de optimizar la gestión de los hospitales, como por ejemplo para modelar la ocupación de las camas de planta o de las unidades de UCI, programar de la manera óptima la utilización de los quirófanos o prever la necesidad de personal médico y de enfermería en cada momento.

En estos casos, se aprovechan los datos históricos que manejan los hospitales acerca de ingresos hospitalarios, intervenciones quirúrgicas, patologías más frecuentes, duración de los ingresos, etc. Para predecir, por ejemplo:

- Cuántos pacientes van a estar ingresados en el hospital cada día y en cada área, para así programar de manera más eficiente el personal sanitario necesario.
- Cuántos pacientes necesitarán ser derivados o entrarán directamente en la UCI, y así poder estar preparado ante cualquier sobrecarga.
- Cuántos días necesitará estar ingresado cada paciente, para hacer una mejor planificación y gestión de los recursos.

Además de la eficiencia en costes, estas aplicaciones redundan es en una mejor calidad en la atención a los pacientes.

La IA no siempre se aplica sobre procedimientos críticos. En este sentido, un algoritmo puede automatizar tareas rutinarias o puramente operativas, permitiendo a los especialistas médicos disponer de más tiempo para atender a los pacientes o para llevar a cabo cualquier otra acción que requiera de su *expertise*.

En todos los casos, por el innegable impacto social que plantean los modelos de IA en el ámbito sanitario, deberemos tener en cuenta el correcto cumplimiento de requisitos como el de "**Bienestar social y ambiental**" precisamente para garantizar ese beneficio para la sociedad de estas soluciones, pero también el de "**Diversidad, no discriminación y equidad**", para asegurar que ningún paciente pueda sufrir ningún tipo de discriminación a la hora de recibir la mejor atención sanitaria disponible.

digitales_

1.-

TRANSPARENCIA & EXPLICABILIDAD

Javier Fernández Castañón

Co-director del área de Salud de OdiseIA

P.- ¿Hasta qué punto la transparencia representa una barrera en el uso médico de la Inteligencia Artificial?

R.- Observamos que muchos profesionales sanitarios desconfían de los sistemas de Inteligencia Artificial. Sólo si comprenden quién los ha programado o cómo llegan a las conclusiones, se supera esa barrera.

P.- ¿Qué cree que preocupa más a los médicos?

R.- La calidad del dato, es decir, que exista una muestra de datos representativa y libre de errores y sesgos de relevancia. Y es que, en medicina, cualquier error podría ser fatal.

P.- ¿Cómo podríamos fomentar la confianza del profesional sanitario?

R.- El primer paso es promover una mayor interrelación entre los ingenieros y los profesionales de la medicina. Un sistema muy avanzado tecnológicamente no servirá de nada si su cliente no lo usa, porque no lo comprende o no confía en que pueda resultar una herramienta útil para su trabajo.

En los ámbitos en los que la Inteligencia Artificial (IA) pueda tener un impacto significativo en los individuos, una IA transparente debería permitir recrear los resultados de cualquier decisión, conociendo la lógica y los datos con los que el modelo fue entrenado, asegurando la trazabilidad y auditabilidad de las decisiones automatizadas. Por lo tanto, las organizaciones necesitan equiparse de herramientas de interpretabilidad y trazabilidad que permitan analizar tanto los datos como los modelos de IA.

La IA explicable es un requisito esencial de los modelos de IA para comprender, confiar y gestionar los sistemas de decisión automatizada. El principio de explicabilidad incluye la transparencia de los elementos pertinentes para un sistema de IA, como son los datos, el modelo y la implicación de sus resultados.

La explicabilidad técnica implica que los procesos utilizados por la IA sean interpretables y permitan la trazabilidad, haciendo posible la revisión del proceso que ha seguido la tecnología para llegar a un determinado resultado. Es importante precisar que este proceso de transparencia no implica la liberación del código fuente.

El principio de explicabilidad es uno de los más relevantes en el sector sanitario, en tanto los resultados que origina la IA tienen impacto sobre las personas. *"En el ámbito de la salud, no hay prueba y error"*, recuerda Javier Fernández Castañón, co-director del área de Salud de OdiseIA.

En este contexto, recomendamos que siempre se revise el proceso de decisión del sistema de IA antes de su implementación en *real*. Hay que tener en cuenta que la explicación del proceso debe ser acorde al nivel técnico de la persona interesada, pudiendo ser una persona no experta en la materia y, en consecuencia, debe ser transmitida de manera clara y precisa. De su lado, todos los profesionales sanitarios deben poseer un conocimiento básico que les permita interaccionar y entender la IA. Conocimientos básicos sobre los fundamentos de la IA, pero también una mejor comprensión de los conceptos matemáticos que hay detrás de esta tecnología, de la procedencia de los datos de salud e, incluso, cuestiones éticas y legales asociadas con el uso de la IA para la salud.

Respecto a la ética de comunicación de la IA, los propios sistemas no deberían presentarse como si fueran "humanos" ante los usuarios. Las personas deben ser conocedoras de que están interactuando con un sistema de Inteligencia Artificial, por lo que se debe mostrar como tal.

Además, se debe ofrecer al usuario la posibilidad de decidir si prefiere interactuar con un sistema de IA o bien con otra persona. Se debería informar a las personas que interaccionen con un sistema de Inteligencia Artificial de las capacidades y limitaciones de éste, es decir, que el usuario conozca de forma exacta el nivel de precisión del sistema.

DESAFÍOS

1.- Uno de los grandes problemas a los que se enfrenta la IA en la sanidad es la adopción de esta tecnología por parte del profesional sanitario. La IA debe ser

comprensible, transparente y explicable para transmitir confianza al profesional sanitario en su uso habitual, sin la necesidad de poseer un conocimiento técnico concreto. El profesional sanitario debe conocer en qué variables o datos se ha apoyado el sistema de Inteligencia Artificial para obtener dicho resultado.

2.- Un sistema de aprendizaje automático basado en algoritmos complejos puede resultar difícil de comprender por algunos usuarios, y difícil de explicar por los desarrolladores.

3.- Se ha comprobado que los profesionales sanitarios prefieren sacrificar un pequeño porcentaje de precisión con tal de aumentar la explicabilidad. Ellos son los que quieren seguir tomando las decisiones y que sea el sistema de IA el que les dé el mayor número de herramientas e información para ello.

4.- En la mayoría de las ocasiones, los usuarios de los sistemas de IA (médicos) no son quienes eligen el proveedor de software adoptado por el centro sanitario en el que trabajan.

5.- Un reto que deben afrontar los ingenieros desarrolladores de código es trabajar juntamente con los profesionales sanitarios para conseguir transmitir una información transparente y de calidad a través del sistema IA. Deben conocer de primera mano qué información es útil y de qué forma presentarla al profesional sanitario.

6.- De su lado, la IA no debe recaer únicamente sobre los desarrolladores del sistema, sino que también los profesionales sanitarios deben recibir una formación homogénea y básica sobre cómo interpretar los datos que el sistema de IA les proporciona.

PROPUESTAS

1.- Promover obligaciones mínimas de transparencia para aplicaciones de menor riesgo: las personas tienen derecho a saber cuándo están interactuando con un sistema de inteligencia artificial.

2.- Definir criterios mínimos de explicabilidad que deban ofrecer los sistemas de IA a los profesionales sanitarios. De forma más concreta, es importante asegurar la trazabilidad sobre:

- Los métodos utilizados para diseñar y desarrollar el sistema algorítmico.
- Los métodos empleados para ensayar y validar el sistema algorítmico.
- Los resultados del sistema algorítmico.

Garantizar la trazabilidad significa que las decisiones ejecutadas por sistemas algorítmicos puedan ser auditadas, evaluadas y explicadas por las personas responsables.

3.- Proporcionar desde las instituciones formación sobre el funcionamiento de la IA en abierto.

4.- Incorporar en el grado de Medicina asignaturas sobre salud digital para actualizar el grado a las demandas contemporáneas de la medicina.

5.- Establecimiento de una responsabilidad legal ante la no transparencia en los sistemas.

Richard Benjamins

Chief AI & Data Strategist. Telefonica

Explicabilidad: la clave de la confianza en la IA

La inteligencia artificial tiene muchas aplicaciones en el sector sanitario. Podemos pensar en aplicaciones que benefician al paciente, que luchan contra enfermedades raras, o que hacen el sistema sanitario más eficiente. También en aplicaciones preventivas, no las que curan a las personas, sino las que consiguen que las personas no enfermen. Dentro de la primera categoría, los beneficios para el paciente, se incluyen aplicaciones de inteligencia artificial para el diagnóstico de enfermedades, para tratamientos personalizados o para el descubrimiento de nuevos medicamentos. También están incluidas las aplicaciones que permiten llevar a un médico virtual a personas que no tienen fácil acceso a los sistemas de salud en, por ejemplo, países en vía de desarrollo con sistemas sanitarios menos avanzados.

Los resultados médicos que generan todas estas aplicaciones tienen mucho impacto en la vida de las personas. **Aunque la fiabilidad pueda ser de un 95%, si se aplica a decenas o centenas de miles de personas, un 5% sigue siendo muchas personas afectadas.**

Por eso, es importante que las conclusiones de los sistemas de inteligencia artificial en el sector sanitario, al menos de aquellas con impacto importante en las personas, sean entendibles y transparentes. **Nunca se debería tomar una decisión sugerida por una máquina que no sea entendible ni transparente para las personas**. Me refiero aquí a todas las personas implicadas en el proceso de IA en el ámbito sanitario. Puede ser el médico que lo usa para llegar a su propia conclusión en un diagnóstico. También puede ser el paciente, que tiene derecho a una explicación sobre por qué el médico piensa que tiene esa enfermedad. Incluso puede ser la organización, por ejemplo, el hospital, que debería ser capaz de entender cómo funciona el modelo de inteligencia artificial en general para relacionar sintomatología y diagnóstico. Hay usuarios adicionales que deben ser capaces de entender cómo funciona el modelo, por ejemplo, un regulador de sistemas médicos, o el regulador de la futura regulación europea de inteligencia artificial.

Los modelos opacos o poco transparentes no deberían aplicarse a decisiones que tienen un impacto significativo en la vida de las personas, en el sector médico o en otros sectores trascendentes. Por ejemplo, para decidir el acceso a servicios públicos como hospitales, colegios o prestaciones sociales, acceso a créditos o hipotecas, o evaluación del rendimiento del trabajo.

Por supuesto, no todas las decisiones son iguales y por eso la transparencia y la explicabilidad son útiles en algunos casos, pero son críticas en otros. No es lo mismo recibir una recomendación de una serie en streaming, que recibir un diagnóstico de una enfermedad grave. En definitiva, en IA no podemos aplicar el *café para todos*. Encontrar el nivel adecuado de transparencia y explicabilidad para las soluciones de IA, en función del uso para el que hayan sido diseñadas, será esencial para lograr la confianza de todos.


digitales_

2.-

PRIVACIDAD

Oliver Smith

Director de Estrategia y Ética de Koa Health

P.- ¿Qué importancia tiene la privacidad en el uso de IA en sanidad?

R.- Es un término claramente ligado a la confianza del paciente, pero no el único. El paciente tiene que confiar en que puede compartir sus datos de salud de forma anónima, sin que, por ejemplo, su jefe, su empresa o su aseguradora pueda jamás acceder a ellos. Esto último preocupa especialmente a la ciudadanía en EE.UU., si bien también en Europa existe una preocupación creciente con respecto a la privacidad de datos de carácter sensible, como los sanitarios y financieros.

El paciente debe entender que sus datos serán tratados de forma agregada y que contribuirán de forma fundamental a que el sistema de IA resulte eficaz. Si la gente no confía en este proceso y dificulta la compartición de datos, esta situación se convierte en una barrera para el desarrollo de algoritmos sanitarios robustos.

P.- ¿Cuál es la clave para generar la confianza del paciente?

R.- Por un lado, hay que ser muy claros -contundentes- con respecto al tratamiento anónimo de sus datos personales. Indirectamente, generar confianza sobre el profesional sanitario acaba impactando en el paciente, aumentando su predisposición a facilitar datos de salud. Y la confianza de los sanitarios está directamente ligada con la calidad de los datos y la transparencia del algoritmo.

La protección de los datos personales es otro de los grandes principios que rigen y condicionan el desarrollo de la IA en el campo sanitario. Así, hay que tener en consideración el carácter particularmente sensible de estos datos.

El Reglamento General de Protección de Datos (GDPR) regula el uso de los datos personales y establece obligaciones para garantizar la privacidad por diseño y por defecto. No obstante, los sistemas de IA amplifican riesgos y conllevan nuevas consideraciones éticas para los equipos tecnológicos. Es clave establecer mecanismos que protejan la Integridad del Dato mitigando riesgos tales como:

- Riesgos de reproducción de patrones erróneos o poco éticos por datos de "mala calidad".
- Riesgos de re-identificación de los individuos.

Por otro lado, se debe asegurar una **adecuada gestión de los datos**, que abarque la calidad e integridad de los datos utilizados, su pertinencia en contraste con el ámbito en el que se desplegarán los sistemas de IA, sus protocolos de acceso y la capacidad de procesar datos sin vulnerar la privacidad de las personas.

La privacidad es un derecho que se ve amenazado por los sistemas de Inteligencia Artificial y que ha de garantizarse a lo largo de todo el proceso de gestión del dato. Esto abarca a los datos que han sido facilitados previamente por cada individuo, así como a los que se van generando en la interacción con el sistema. Los datos que se extraen del comportamiento humano pueden conducir a que el sistema de Inteligencia Artificial los clasifique en función de su orientación sexual, su edad, género u opiniones políticas y religiosas. Por este motivo, para que los individuos puedan confiar en el proceso de recopilación de datos, es preciso garantizar que la información recabada sobre ellos no se utilizará para una discriminación injusta o ilegal.

Otro aspecto esencial es la **calidad e integridad de los datos.** El proceso de recopilación de información puede contener sesgos sociales, imprecisiones o errores. Particularmente en ámbitos como el sanitario, resulta esencial corregir estos errores y garantizar la integridad de los sistemas, antes incluso de empezar a usarlos. El daño potencial que puede recibir un sistema de IA con la introducción de datos erróneos es muy elevado, sobre todo si se trata de sistemas con capacidad de autoaprendizaje.

Por último, es fundamental, sobre todo si se trata de datos personales, **controlar quién puede acceder y gestionar los datos a lo largo del ciclo de vida de la IA**. En estos protocolos debería describirse quién puede acceder a los datos y bajo qué circunstancias. Así, solamente debería tener acceso a los datos personales el personal debidamente cualificado, poseedor de las competencias adecuadas y que necesite acceder a la información pertinente.

En todos los casos, cualquier recopilación de datos personales debe venir precedida de un consentimiento informado por parte del ciudadano. Incluso, para su uso por parte de Administraciones públicas. Un ejemplo que ha sido largamente debatido desde marzo de 2020 es el rastreo de la población para el seguimiento de epidemias y prevención de crisis futuras. El mismo principio de privacidad debería prevalecer también en la utilización de datos procedentes de la

historia clínica para reforzar ensayos clínicos o para estudios médicos.

La eficacia de cualquier sistema de IA está directamente ligada a la confianza que los pacientes y los profesionales sanitarios depositen en él. A este respecto, la información proporcionada al paciente sobre el tratamiento anónimo de sus datos personales ha de ser absolutamente clara. La confianza de los pacientes será determinante para obtener muestras suficientemente representativas de información, que permitan al algoritmo funcionar correctamente.

En España, la información al paciente constituye no sólo una recomendación, sino una obligación legal. Desde la publicación del Reglamento General de Protección de Datos (RGPD) esta información es considerada como "datos especialmente protegidos". Esto implica que los datos sólo se pueden tratar bajo consentimiento expreso y por escrito del afectado. Además, las entidades que traten estas categorías especiales de datos, sobre todo si lo hacen a gran escala, han de realizar una evaluación de impacto y contar con un Delegado de Protección de Datos. Este marco normativo es complejo y requiere del pertinente asesoramiento legal profesional.

PSEUDO-ANONIMIZACIÓN

La desidentificación evita la conexión de identificadores personales con la información. La anonimización de los datos personales es una subcategoría de la desidentificación, donde se eliminan los identificadores personales directos e indirectos, y se utilizan salvaguardas técnicas para garantizar un riesgo cero de re-identificación, mientras que los datos desidentificados se pueden volver a identificar mediante el uso de una clave.

Por su parte, la pseudo-anonimización permite que los datos no puedan atribuirse a un sujeto específico sin el uso de información adicional, manteniéndose esa información adicional por separado.

La pseudo-anonimización se presenta, en este contexto, como la alternativa más segura para la mayor parte de contextos. En cambio, la desidentificación puede comprometerse, en muchos casos, con una simple triangulación de información.

La anonimización, por su parte, puede restar robustez y eficacia a un sistema de IA aplicado para la búsqueda de diagnósticos o tratamientos para patologías complejas, donde cada persona requiere un estudio individualizado, en función de su perfil sociodemográfico, sus antecedentes familiares o sus condiciones médicas preexistentes.

DESAFÍOS

1.- Asegurar que los datos que maneja la IA sean tratados de forma responsable y únicamente para los usos para los cuales fueron recabados. Particularmente, cuando se trate de colectivos vulnerables, como menores.

2.- Desarrollar un mecanismo ágil, eficiente y transparente, riguroso a la par que sencillo, que permita al usuario dar su consentimiento y entender el alcance de éste.

3.- Conseguir el máximo aprovechamiento de los datos personales con los que cuentan los organismos sanitarios, manteniendo siempre la privacidad de éstos.

4.- El derecho a la privacidad prevalece sobre otros derechos que pudieran ser chocar con éste. En este sentido, una limitación en la privacidad de los datos personales podría implicar restricciones en la capacidad de los ciudadanos de ejercer de forma completa otros derechos civiles, políticos, sociales o económicos.

5.- A medida que proliferan las fuentes de datos de carácter sanitario -registros médicos electrónicos, imágenes radiológicas, pero también tests genéticos ofertados por agentes privados, apps para deportistas, relojes inteligentes…-, crecen también los potenciales focos de ciberataques.

6.- Falta de procedimientos estandarizados para la pseudo-anonimización de los datos y la protección de aquella información complementaria que permitiría triangular o desanonimizar los datos.

PROPUESTAS

1.- Otorgar la formación y conocimientos necesarios a las empresas/centros sanitarios para asegurar el cumplimiento de la normativa en protección datos, evitando confusiones o actuaciones ilícitas fuera del margen legal.

2.- Crear mecanismos que aseguren que el desconocimiento de la ley no desemboca en una actividad ilegal.

3.- Incorporar mecanismos robustos de pseudo-anonimización de los datos que garanticen la privacidad de éstos a la vez que pueden ser utilizados para fines de investigación de nuevos medicamentos, tratamiento de enfermedades, etcétera.

4.- Desarrollar planes de contingencia en caso de filtración, robo, pérdida o re-identificación de los datos.

BEST PRACTICE 1 - PROYECTO COHORTE

El **Gobierno de Cantabria**, a través de la Dirección General de Salud Pública, ha puesto en marcha un proyecto denominado “Cohorte Cantabria”, por el que solicita a los ciudadanos que cedan voluntariamente muestras y datos de salud, de forma vitalicia, con el objetivo de impulsar la investigación biomédica y avanzar hacia una medicina de precisión. El público objetivo es toda la ciudadanía de entre 40 y 70 años, unas 50.000 personas, equivalente al 18% de la población de esta Comunidad Autónoma.

El proyecto se puso en marcha a mediados de 2021 y cuenta ya con más de 7.000 voluntarios. La investigación, que conjugará datos clínicos y biológicos con datos sociales, ayudará a conocer y comprender las causas y el pronóstico de distintas patologías agudas y crónicas que afectan a la población de Cantabria. La base de datos incorporará información sobre factores determinantes de salud como los estilos de vida, aspectos socioeconómicos, académicos, demográficos, analíticos y de enfermedad de la población. Se pretende analizar la evolución de estas cifras en el tiempo y determinar su influencia en el desarrollo de enfermedades.

Marta Pastor Villalobos

AI Strategy Lead Analyst @NTT DATA AI Center of Excellence

La privacidad: una cuestión de voluntad y esfuerzo de todos

El auge de la IA en el sector sanitario y la aceleración de su adopción debido a la crisis del COVID-19 hacen que los retos y peligros relacionados con la privacidad de los datos sean ya una realidad. Si tenemos en cuenta además el alto grado de sensibilidad de los datos con los que tratamos, nos encontramos con un problema actual de gran importancia para el conjunto de la sociedad.

La recopilación masiva de datos y su difusión pública justificada por la emergencia sanitaria ya ha provocado situaciones de discriminación como ocurrió en Corea en 2020. La publicación de nombres de bares y clubs de temática LGBT como focos de contagio, debido a que una persona con positiva en coronavirus visitó varios de ellos, llevó a la culpabilización del colectivo por el incremento de casos y aumentó el temor de las personas del colectivo a ser reconocidos por su orientación sexual.

La privacidad de los datos médicos de los ciudadanos también se ven expuestos de manera menos explícita mediante nuevos dispositivos y servicios digitales como son los ***wearables* o pulseras de actividad**. Dispositivos capaces de monitorizar nuestra actividad diaria, ritmo cardíaco, calidad del sueño con el objetivo de empoderarnos a tener un mayor control de nuestra salud, pero **sin determinación clara de quién es el dueño legal de dichos datos ni de qué hacen o pueden hacer las compañías tecnológicas con ellos**.

Estos ejemplos ponen de relieve la necesidad de un mayor conocimiento y control sobre este tipo de información. Es primordial un **esfuerzo conjunto**: partiendo de tener una legislación actualizada al estado presente de las tecnologías y que proteja la privacidad de los datos y la soberanía de los ciudadanos sobre ellos, continuando por realizar **evaluaciones en todos los proyectos de IA** para prevenir posibles peligros y garantizar el uso responsable de la información, y culminando con una mayor concienciación social sobre los servicios y productos que utilizamos.

Desde el punto de vista tecnológico **existen herramientas capaces de mitigar parte de estos riesgos**. Uniendo capacidades de IA con criptografía y tecnologías distribuidas como Blockchain se puede crear un nuevo modelo de Identidad Digital soberana que empodere a las personas haciéndolas dueñas y gestoras de sus datos y credenciales. A nivel criptográfico se pueden crear pruebas de conocimiento cero (en inglés *Zero Knowledge proof*) que permitan la verificación de cierta información sin revelar datos personales. Y a nivel de soluciones de IA existen nuevos modelos de entrenamiento como el *Distributed Learning model* que garantiza la soberanía y control de los datos en todo momento.

Existen, pues, soluciones y alternativas para solventar los riesgos de la privacidad; solo es necesario contar con voluntad y esfuerzo por parte de todos.


digitales_

3.-

CALIDAD DEL DATO & NO DISCRIMINACIÓN

Gemma Galdon

Fundadora de Eticas Research & Consulting

P.- La IA existe desde hace décadas y, sin embargo, la cuestión sobre su dimensión ética es bastante reciente. ¿Por qué?

R.- Durante años, los algoritmos se usaban para cuestiones más o menos triviales. Su aplicación en ámbitos sensibles, como la sanidad, es más reciente.

P.- ¿Cuál es la mayor barrera para una buena gobernanza de los datos?

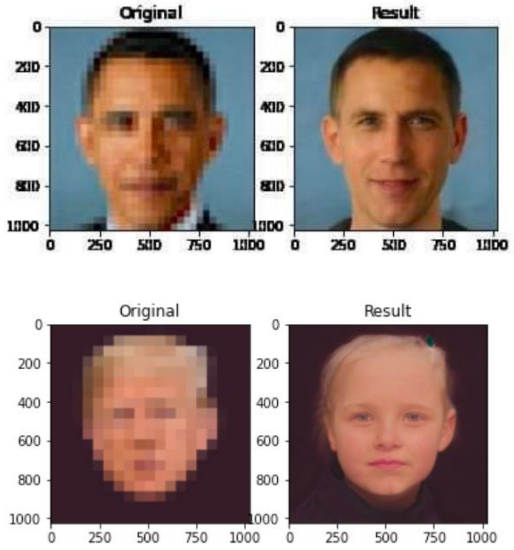
R.- Requiere planificación y una política interna. Hay veces que la política tecnológica se limita a comprar un algoritmo, pero se olvida todo lo que viene después. Requiere, por tanto, establecer mecanismos de gobernanza de los datos. Y las consideraciones para definir esos mecanismos dependen del contexto. Cada uso de un sistema de IA puede requerir diferentes niveles de anonimización y de control, por ejemplo.

El principio de no discriminación, en IA, consiste en procurar una calidad del dato que resulte compatible y respetuosa con la diversidad y la equidad.

La IA tiene el potencial de ayudar a las personas a tomar decisiones más justas, pero solo si diseñamos sistemas responsables, evaluando sesgos a lo largo del ciclo de vida del algoritmo. El diseño responsable de la IA tiene como objetivo evitar que las decisiones de los algoritmos perpetúen y amplifiquen los sesgos y desigualdades sociales ya presentes en nuestra cultura, evitando una discriminación aún mayor. Uno de los mayores retos a los cuales se enfrentan los científicos de datos al corregir sesgos es la potencial pérdida de calidad del sistema de IA, ya que esta corrección puede implicar una menor precisión del sistema.

Por ello, se presenta esencial evitar sesgos injustos en los sistemas de Inteligencia Artificial. Los conjuntos de datos que utilizan los sistemas de IA pueden presentar sesgos históricos inadvertidos, lagunas o modelos de gestión incorrectos. Esto podría dar lugar a prejuicios y discriminación indirectos e involuntarios contra determinados grupos o personas, lo que podría agravar los estereotipos y la marginación. Peor aún, **en el ámbito sanitario, la maximización de sesgos preexistentes puede acarrear riesgos sobre la seguridad de los pacientes, la ciberseguridad o el entorno**.

Asimismo, un reciente informe de la Organización Mundial de la Salud (OMS) advierte de que los sistemas capacitados con datos recopilados de personas de nivel socioeconómico medio y alto pueden no funcionar bien para las personas de entornos más desfavorecidos[4]. En este sentido, la OMS postula que los sistemas de IA deben diseñarse cuidadosamente para reflejar la diversidad de entornos socioeconómicos y de atención sanitaria.



Fuente: Twitter / @Chicken3gg / Moritz Klack

La explotación intencionada de los sesgos de los consumidores o la competencia desleal también pueden provocar situaciones perjudiciales, como la homogeneización de los precios mediante la colusión o la falta de transparencia del mercado. Siempre que sea posible, los sesgos identificables y discriminatorios deberían eliminarse en la fase de recopilación de la información.

Los propios métodos de desarrollo de los sistemas de IA (como puede ser la programación de algoritmos) también pueden presentar sesgos injustos. Esto se puede combatir mediante procesos de supervisión que permitan analizar y abordar el propósito, las restricciones, los

[4] https://www.who.int/news/item/28-06-2021-who-issues-first-global-report-on-ai-in-health-and-six-guiding-principles-for-its-design-and-use

requisitos y las decisiones de un modo claro y transparente. Además, promover la **diversidad** en equipos que desarrollen IA, incluyendo a personas procedentes de diversos contextos, culturas y disciplinas, puede garantizar una mayor inclusión y representatividad en los sistemas de IA.

En el ámbito específico de las relaciones entre organizaciones y consumidores, los sistemas deberían estar centrados en el usuario y diseñarse de un modo que permitan que todas las personas utilicen los productos o servicios de IA con independencia de su edad, género, capacidades o características. En este sentido, la **accesibilidad** de esta tecnología para las personas con discapacidad, que están presentes en todos los grupos sociales, reviste una importancia particular.

Los sistemas de IA deben ser adaptables y tener en cuenta los principios del 'Diseño Universal' para servir al mayor número posible de usuarios, observando las normas de accesibilidad pertinentes. Esto permitirá un acceso equitativo y una participación activa de todas las personas en las actividades humanas informatizadas existentes y emergentes, así como en lo que atañe a las tecnologías asistenciales.

Con el propósito de desarrollar sistemas de IA fiables, se recomienda consultar a todas las partes interesadas que puedan ser afectadas de manera directa o indirecta por el sistema durante todo su ciclo de vida. Conviene pedir opiniones periódicamente incluso después del despliegue de los sistemas de IA y establecer mecanismos para la participación de las partes interesadas a largo plazo, por ejemplo, garantizando la información, consulta y participación de los trabajadores a lo largo de todo el proceso de implantación de este tipo de sistemas en las organizaciones.

Ahora bien, **¿por qué si la IA es tan potente, su adopción no lo es?** ¿Por qué hay sectores donde su uso está muy implementado y otros no tanto? ¿Estamos sobrevalorando las capacidades de la IA? Observamos que **muchos de los fallos de la IA vienen precedidos de errores en la definición del proyecto y sus objetivos**, en ocasiones, basados en una imagen no realista sobre las potencialidades de estas tecnologías. Un sistema de IA puede no tener mucha efectividad ante enfermedades raras o fenómenos nuevos, como epidemias desconocidas y sobrevenidas, si la disponibilidad de información es menor.

BEST PRACTICE 2 – AMILOIDOSIS CARDÍACA

La amiloidosis cardíaca es una enfermedad rara de difícil diagnóstico, debido a que sus síntomas suelen confundirse con otros padecimientos propios de la edad. **Sopra Steria**, la **Fundación San Juan de Dios** y el **Hospital San Juan de Dios de León** desarrollaron un proyecto de investigación para mejorar la predicción de esta enfermedad usando IA. Sobre la base de 10 años de historias clínicas anonimizadas de casi 16.000 pacientes, se realizó un estudio sobre una muestra de personas con fallo cardíaco, un diagnóstico relevante para la amiloidosis, aplicando mejoras en el procesamiento de los datos y en los modelos analíticos mediante Big Data y algoritmos de Machine Learning, logrando detectar 50 casos no diagnosticados.

En definitiva, **en la recogida de datos para los sistemas de IA importa tanto la cantidad como la calidad** de éstos. Una calidad que debe ser medible o auditable, superando dificultades como la dificultad del propio proceso de recogida de información por parte del personal administrativo y sanitario, o la dificultad de procesar datos no estructurados, combinados con la suficiente información sobre el entorno. Estos desafíos sobre la propia calidad de los datos corroboran la necesidad de entender la relación persona-máquina como un binomio donde siempre prevalezca la supervisión humana (ver Capítulo 5).

DESAFÍOS

1.- El proceso de recogida de datos puede ser complejo y costoso.

2.- Conseguir que cualquier persona, independientemente de su condición social, capacidades u otros factores, pueda hacer uso de soluciones basadas en IA que ayuden en tareas como el autodiagnóstico o la telemedicina.

3.- Evitar cualquier tipo de comportamiento sesgado en los sistemas de IA que tenga su origen en los propios datos utilizados para su entrenamiento y/o en el proceso de construcción de dicho sistema.

4.- La IA debe tener aprendidos los principios de derechos humanos, así como de justicia. Que no trate a los pacientes como si de meros números se trataran. Que el porcentaje de éxito, supervivencia o coste no sean factores que permitan a la IA libremente desechar pacientes.

5.- La IA debe ser capaz de no discriminar a un paciente por su condición de salud o edad de cara a recibir un tratamiento, posicionarlo en una lista de espera o descartarlo por alto coste del tratamiento.

6.- Para identificar e, idealmente, anticipar los posibles efectos perjudiciales de la IA sobre las desigualdades, la observación o estudio del impacto de los algoritmos debería tener una continuidad en el tiempo.

7.- La ética debe estar integrada en todas las etapas de diseño, desarrollo e implementación de sistemas de IA.

8.- La IA maximiza sesgos previos presentes en la cultura y en la sociedad. Erradicar dichos sesgos implicaría un debate que trasciende el ámbito sanitario.

PROPUESTAS

1.- Promover iniciativas de formación y capacitación a toda la población, especialmente la más vulnerable, sobre el uso de herramientas basadas en IA para el ámbito sanitario.

2.- Requerir a los desarrolladores de sistemas de IA la inclusión de mecanismos de control, incluyendo la intervención humana, que garanticen la no presencia de sesgo de ningún tipo en sus soluciones.

3.- Acelerar la interoperabilidad de sistemas, asegurando la privacidad y la calidad de los datos, para ampliar el tamaño de la muestra de información que vaya a ser procesada. En esta línea, España está trabajando ya en la creación de un *data lake* (repositorio único) de datos sanitarios.

4.- Incorporar a las acciones de formación y explicabilidad dirigidas a personal sanitario advertencias sobre los posibles sesgos de la información que manejan.

Cristina Aranda

Cofundadora de Big Onion. Confundadora de Mujeres Tech. Spain AI y ELLIS Alicante

Lo esencial es invisible a los ojos… ¿Y a las máquinas?

Retomando la lectura de *El Principito* nos adentramos en cuestiones inherentemente humanas: la ternura, la amistad, la imaginación… y la solidaridad. "No se ve bien sino con el corazón", le dice el zorro a protagonista de este cuento poético, algo que las máquinas no poseen, por ahora.

Vivimos en plena revolución industrial liderada originada en gran medida por la generación y uso masivo de datos y la inteligencia artificial (IA), quien se encarga de darles sentido, extraer patrones, tomar decisiones y automatizar procesos, entre otras tareas, donde las personas somos las principales protagonistas. En la Sanidad, son muchos los procesos donde interviene un algoritmo o una IA para determinar una enfermedad, el orden de operaciones o incluso las operaciones en remoto.

La tecnología es un medio para hacer que nuestro día a día sea mejor. Sin embargo, las personas que trabajamos con esta tecnología no disponemos de un código ético como sí cuentan las personas que trabajan en el sector sanitario, sobre todo para saber cómo *hackear* nuestros sesgos.

La materia prima de la IA son los datos. Sin datos no hay IA. Ahora bien, estos datos los generamos y los manipulamos las personas. Pero también tenemos sesgos a la hora de escoger un determinado modelo para entrenar a la máquina. Esto lo explican muy bien Ricardo Baeza y Karma Peiró en su ilustrativo artículo "¿Es posible acabar con los sesgos de los algoritmos?". **Además de aplicar sesgos a la hora de usar la tecnología, también tenemos sesgos que proceden de nuestra educación y cultura**. Son muchos los ejemplos de productos y servicios de IA que podría exponer donde se marginan a mujeres, personas negras, asiáticas, musulmanas… sencillamente porque en el equipo de desarrollo no hay diversidad. En el documental "Coded Bias" se exponen y explican todos estos ejemplos y su consiguiente impacto negativo en la sociedad. **Una de las explicaciones es la falta de diversidad de los equipos de desarrollo de IA**, donde el 80% están compuestos por hombres, blancos, judeo-cristianos, de mediana edad, heterosexuales y sin discapacidad. Un caso contrario, ejemplo de la colaboración entre equipos diversos en identidad y formación, es el caso de STOP, un algoritmo para detectar la depresión y prevenir el suicidio liderado por la Dra. Ana Freire de la UPF Barcelona School of Management. Este algoritmo ha conseguido incrementar un 60% de llamadas al teléfono de la esperanza.

En el terreno de la medicina o la sanidad, estos sesgos son visibilizados por personas como la Dra. Carme Valls, doctora especializada en endocrinología y autora de "Mujeres invisibles para la medicina", quien explica en su libro, entre otras muchas cosas, por qué cuando un hombre va a urgencias con un dolor de pecho se le hace directamente un electrocardiograma y con los mismos síntomas a las mujeres se nos receta un ansiolítico para la ansiedad. Estos datos luego se usan para entrenar a una máquina con el consiguiente impacto negativo, en especial, al 50% de la población, a las mujeres.

Así pues, **urge detenernos, analizar, "filosofar" y llegar a un acuerdo entre todas las personas que trabajamos con datos, IA y salud sobre cómo debemos trabajar los datos, los modelos, los algoritmos y cualquier aspecto relacionado con la IA** para hacer que lo esencial no sea visible solo para los ojos, sino también **sentible** para las personas y su bienestar.


20

4.-

ROBUSTEZ

Nerea Luis

Ingeniera en IA en Sngular. Doctora cum laude en inteligencia artificial

P.- ¿Qué significa para ti la robustez en un sistema IA?

R.- La entiendo como la solidez técnica de los sistemas de IA. La robustez se consigue a través de una supervisión y evaluación constante.

P.- Si tuviéramos que llevarnos ese concepto de robustez al ámbito sanitario…

R.- Diría que a través de la robustez de los sistemas de IA transmitimos seguridad al ciudadano. Un sistema de IA robusto es aquel que ha sido entrenado y evaluado frente a los procesos y a las situaciones que va a enfrentar. A posteriori, debemos continuar evaluando a los sistemas de IA frente a una batería de pruebas que avalen su correcto funcionamiento. Por ejemplo, en el campo de sanidad, su reto será mantenerse robusto manejando y gestionando datos atípicos que tan presentes están en el ámbito sanitario.
Es un trabajo sostenido en el tiempo.

La solidez técnica es un requisito fundamental para configurar sistemas de IA robustos, que se desarrollen con un enfoque preventivo de riesgos, al comportarse de la manera esperada. En función de la criticidad de la decisión tomada por el sistema de IA, es necesario asegurar la exactitud y reproducibilidad de los resultados, así como la resiliencia ante potenciales ataques y manipulaciones.

Para ello, es necesario que los sistemas de Inteligencia Artificial se desarrollen con un enfoque preventivo en relación con los riesgos, de modo que se comporten siempre según lo esperado y minimicen los daños involuntarios e imprevistos, evitando asimismo causar daños inaceptables. Lo anterior debería aplicarse también a los cambios potenciales en entorno operativo o la presencia de otros agentes (humanos y artificiales) que puedan interactuar con el sistema de manera contenciosa. Además, debería garantizarse la integridad física y mental de los seres humanos.

Los sistemas de IA, como los de software, **deben estar protegidos frente a las vulnerabilidades** que puedan permitir su explotación por parte de agentes malintencionados como, por ejemplo, ataques cibernéticos. Ataques que pueden ir dirigidos contra los datos (envenenamiento de los datos), el modelo (fallo del modelo) o la infraestructura informática subyacente, tanto el software como el hardware. En el caso de que un sistema de Inteligencia Artificial sea objeto de un ataque por agentes malintencionados, se podrían alterar los datos y el comportamiento del sistema, de modo que este adopte decisiones diferentes o, sencillamente, se desconecte.

Los sistemas y los datos también pueden corromperse debido a intenciones maliciosas o por verse expuestos a situaciones inesperadas. Unos procesos de seguridad insuficientes también pueden dar lugar a decisiones erróneas o incluso a daños físicos, Para que los sistemas de IA se consideren seguros, es preciso tener en cuenta las posibles aplicaciones imprevistas de la IA (aplicaciones que puedan utilizarse para fines diferentes, por ejemplo) así como el abuso potencial de un sistema de IA por parte de agentes malintencionados; también se deberán adoptar medidas para prevenir y mitigar esos riesgos.

Los sistemas de IA deberían contar con **salvaguardias que posibiliten un plan de repliegue en el caso de que surjan problemas**. Esto puede significar que los sistemas de IA pasen de un procedimiento basado en estadísticas a otro basado en normas, o que soliciten la intervención de un operador humano antes de proseguir con sus actuaciones. Es preciso garantizar que el sistema se comportará de acuerdo con lo que se espera de él sin causar daños a los seres vivos ni al medio ambiente. Esto incluye la minimización de las consecuencias y errores imprevistos. Además, se deberían establecer procesos dirigidos a aclarar y evaluar los posibles riesgos asociados con el uso de sistemas de IA en los diversos ámbitos de aplicación.

El nivel de las medidas de seguridad requeridas depende de la magnitud del riesgo que plantee un sistema de IA, que a su vez depende de las capacidades del sistema. Cuando se prevea que el proceso de desarrollo o el propio sistema planteará riesgos particularmente altos, es crucial desarrollar y probar medidas de seguridad de forma proactiva.

La precisión está relacionada con la capacidad de un sistema de IA para realizar juicios correctos —como, por ejemplo, clasificar correctamente información en las categorías adecuadas—, o con su capacidad para efectuar predicciones, formular recomendaciones o tomar decisiones correctas basándose en datos o modelos. Un proceso de desarrollo y evaluación explícito y correctamente diseñado puede respaldar, mitigar y corregir los riesgos imprevistos asociados a predicciones incorrectas. Cuando no sea posible evitar este tipo de predicciones, es importante que el sistema pueda indicar la probabilidad de que se produzcan esos errores. Un alto nivel de precisión resulta particularmente crucial en situaciones en que un sistema de IA afecte de manera directa a la vida humana.

¿Chihuahua o magdalena?

Fuente: @TEENYBISCUIT

Para la fiabilidad es esencial que los resultados de los sistemas de IA sean también reproducibles. Un sistema de IA fiable es aquel que funciona adecuadamente con un conjunto de información y en diversas situaciones. Esto es necesario para evaluar un sistema de IA y evitar que provoque daños involuntarios. La reproducibilidad describe si un experimento con IA muestra el mismo comportamiento cuando se repite varias veces en las mismas condiciones. Esto permite a los científicos y responsables políticos describir con exactitud lo que hacen los sistemas de IA. Los archivos de replicación pueden facilitar el proceso de ensayo y reproducción de comportamientos.

Conviene incidir en el riesgo de aplicar técnicas no probadas durante una situación de emergencia. En casos como la pandemia del Covid-19, sobrevenida de forma inesperada, no existían datos históricos a partir de los que poder efectuar predicciones de forma efectiva. Un sistema de IA debería cumplir con los estándares de validez científica y precisión que se aplican actualmente al resto de las tecnologías médica o, en su defecto, una estimación del valor de credibilidad que se puede asignar a las recomendaciones que haga el sistema.

DESAFÍOS

1.- Se debe conseguir una distribución de datos homogénea y una obtención de estos de calidad. Con esto se consigue que el sistema de IA identifique fácilmente correlaciones y patrones. Relevante también, la utilización de los datos correctos, que la muestra sea representativa de la realidad.

2.- Conseguir que las Administraciones otorguen unos datos públicos de calidad, evitando disparidades. Por ejemplo, en España, el sistema sanitario tiene la peculiaridad de estar dividido en 17 Comunidades Autónomas, lo que

complica crear bases de datos eficientes que aprovechen todos los datos disponibles.

3.- No existe una única historia clínica para cada persona. Así, a la descentralización del sistema público de salud en España, han de sumarse los historiales que manejan, por ejemplo, las mutuas de trabajo, los seguros de salud, etcétera.

4.- Asegurar que los sistemas de IA cumplen unos niveles mínimos de precisión para el uso para o por parte de la población.

5.- Conseguir que los sistemas de IA sean capaces de ofrecer respuestas en situaciones reales (lo que conocemos como "entornos productivos") con los mismos niveles de precisión y fiabilidad que en las etapas de desarrollo y pruebas.

PROPUESTAS

1.- Definir unos niveles mínimos de fiabilidad y precisión para cada tipo de aplicación de los sistemas de IA que se deban cumplir antes de ponerlos en producción, en función de la sensibilidad del caso de uso.

2.- Promover la implantación de un sistema de protección o de almacenamiento (nube) que permita crear copias de seguridad y proteger los datos de pacientes ante cualquier tipo de inconveniente, ya sea apagón, error o destrucción del sistema, ataques de seguridad.

3.- Promover un sistema de certificación para software de IA en sanidad a nivel nacional, basado en criterios objetivos de calidad de las auditorías algorítmicas, e impulsar esa certificación a nivel europeo.

4.- No todos los modelos algorítmicos ofrecen los mismos resultados en los distintos casos de uso. El *expertise* en el ámbito *health* (y la experiencia previa de otras investigaciones y organizaciones) será esencial para saber escoger entre Machine Learning y Deep Learning y, dentro de la primera categoría, entre aprendizaje supervisado, aprendizaje no supervisado y aprendizaje reforzado.

BEST PRACTICE 3 – JANO

El **Gobierno de Cantabria**, a través de la Dirección General de Transformación Digital de la Conserjería de Sanidad, ha puesto en marcha varias herramientas digitales dirigidas a profesionales sanitarios y a ciudadanos durante la pandemia del COVID-19. Queremos aquí prestar atención a 'Jano', un robot capaz de reconocer el lenguaje natural, que ha realizado más de un millón de llamadas para concertar citas de vacunación. 'Jano' puede realizar 200 llamadas simultáneas, lo que claramente contribuye a descargar de tareas al personal administrativo. La herramienta aprende con el tiempo y ha podido ponerse en producción antes de comprobar o de alcanzar una robustez máxima, dado que su uso acarrea menores riesgos que un sistema de IA de uso clínico, por ejemplo. En cualquier caso, sus creadores estiman que 'Jano' *entiende* al 97,7% de las personas con las que interactúa. 'Jano' es obra de la empresa local **Idrus** y está basado en tecnología de Google.

CONSIDERACIONES SOBRE LA AUDITORÍA ALGORÍTMICA

Siempre que sea posible, es conveniente auditar los algoritmos, prestando especial atención sobre su robustez, su trazabilidad y su valor ético. La práctica de la auditoría algorítmica se realiza sobre un sistema ya implementado, si bien sus fundamentos pueden tenerse en cuenta desde la fase de desarrollo de las tecnologías.

Actualmente no existe una metodología de auditoría algorítmica. Es un espacio aún incipiente, donde existe una experiencia limitada. La auditoría más extendida es puramente técnica y puede realizarse incluso empleando software.

Idealmente, un ejercicio de auditoría en un campo sensible como la sanidad debería integrar también un análisis más amplio, incluyendo una valoración de la implementación. Y, dentro de esta última, se debe analizar cómo se ha explicado a los profesionales el funcionamiento de la IA. Dicho de otro modo: la robustez de un algoritmo, por sí sola, no garantiza su solidez. El usuario [en este caso, el médico] debe haber recibido la formación necesaria e instrucciones claras sobre el nivel de supervisión humana requerido. En última instancia, si no existe un sistema de gobernanza de los datos suficientemente robusto y si el usuario no confía en la IA, ésta no valdrá de nada.

Desde **Eticas Research & Consulting**, recomiendan asimismo que los algoritmos ofrezcan no sólo una recomendación -o variedad de recomendaciones- al usuario, sino también el valor que tienen éstas, en función de la fiabilidad estimada de las mismas. El resultado de este ejercicio de evaluación no suele tener un valor numérico, sino que consta de una serie de recomendaciones de mejora.

Juan Carlos Sánchez Rosado

IBM Health Industry Leader

Rafael Fernández Millán

IBM Technology Partner Architect

La robustez en todo el ciclo de vida de la IA

La aplicación de inteligencia artificial en el sector Salud presenta importantes retos para garantizar la robustez de los modelos a lo largo de su ciclo de vida, entendida como la consistencia de su precisión en el tiempo. Los algoritmos empleados en el diagnóstico, prevención, seguimiento, predicción, pronóstico, tratamiento o alivio de una enfermedad, lesión o discapacidad están sujetos al nuevo **Reglamento Europeo de Productos Sanitarios**, que plantea amplios requerimientos en cuanto a la evaluación clínica del producto, la gestión del riesgo, el sistema de gestión de la calidad, el seguimiento posterior a la implantación, la documentación técnica y la responsabilidad por productos defectuosos.

Durante el periodo de preparación del modelo, los investigadores utilizan unos datos que representan la realidad contrastada y sobre los cuales se ha entrenado el modelo o, dicho de un modo muy simple, **el punto de partida del modelo es un conjunto concreto de datos de entrenamiento**. Mediante un conjunto de validaciones de los resultados de los modelos (precisión, falta de sesgo, matrices de confusión, trazabilidad del dato, etc.), se verifica su idoneidad y se sustenta la aprobación para su uso en el entorno sanitario bajo unos márgenes y condiciones. Pero estos resultados no garantizan que, una vez puesto el modelo en producción, no se produzcan cambios en las precisiones del modelo por variaciones en la composición de la población que hace uso del modelo. En consecuencia, deben existir unos mecanismos de validación durante la operación.

A su vez, durante la operación y de modo continuo, el proveedor del modelo debe poder evaluar si se están produciendo variaciones significativas en los modelos sin necesidad de validación clínica. Esta supervisión automática debe permitir: 1) detectar sesgos sobre las variables sensibles que se determinen para el caso de uso del modelo (por ejemplo: edad, sexo, enfermedades previas, etc) fijando ventanas de revisión y umbrales de alarmado, y 2) detectar cambios significativos de población frente a la de entrenamiento que afecten a la precisión.

Por otra parte, aunque la validación clínica *a posteriori* de las predicciones por los facultativos permite realizar un análisis sobre la calidad del modelo utilizando los indicadores aplicados en el entrenamiento (F1, recall o ROC), existen retos significativos para considerarlos fuente única de supervisión. Estos retos incluyen el retraso en la validación del error del modelo (horas, días o semanas después de la ejecución), que la validación es potencialmente incompleta pues algunas situaciones son susceptibles de ser menos informadas (p.ej de falsos positivos en un predictor de sepsis), o que la validación depende de la situación y criterio del propio facultativo revisor.

En conclusión, el despliegue de modelos de inteligencia artificial en el entorno sanitario requiere de un **cuidado entorno operativo** con metodologías y herramientas automatizadas y distribuidas, de apoyo a la gestión de la precisión de los modelos, en todo su ciclo de vida que mitigue los riesgos inherentes a la inteligencia artificial.


digitales_

5.-

AUTONOMÍA & SUPERVISIÓN

Javier Mendoza

Consultor en Savana Medical

P.- ¿Cuál crees que es el mayor reto de la IA en la sanidad?

R.- Sin duda, la aceptación por parte del profesional sanitario y el conocimiento [insuficiente, en muchos casos] que éstos poseen sobre el funcionamiento del sistema de IA.

P.- ¿Qué tipo de relación médico–IA debe existir para superar esta barrera?

R.- El médico debe entender que la última decisión siempre será suya; la supervisión humana debe prevalecer. Los algoritmos deben tener la suficiente autonomía como para generar resultados y el profesional sanitario la suficiente jerarquía como para controlar la toma de decisiones. Los estudios demuestran que una combinación de las partes siempre resulta en mejores resultados en comparación al trabajo individual.

Resulta tentador recordar las tres leyes de la robótica que postuló Isaac Asimov en su libro *Círculo Vicioso*, en 1942:

1.- *Un robot no hará daño a un ser humano o, por inacción, permitirá que un ser humano sufra daño.*
2.- *Un robot debe obedecer las órdenes dadas por los seres humanos excepto si estas órdenes entrasen en conflicto con la 1ª ley.*
3.- *Un robot debe proteger su propia existencia en la medida en que esta protección no entre en conflicto con la 1ª o la 2ª Ley.*

Siempre, y de forma particular en el ámbito de la salud, la IA responsable debe de estar centrada en el ser humano. Los humanos cumplimos un doble rol, como sujeto a través de los datos y luego como participantes activos, estando en control de los sistemas de IA. Dentro de todos los sistemas de la IA, es importante trasladar este principio a un nuevo modelo colaborativo "humano-máquina", aumentando las capacidades de supervisión resultando en una mayor profesionalización de los empleados. Una implementación responsable de la IA debería de poner el foco en usos donde la IA automatice aquellas tareas que generen un menor valor para las organizaciones y el personal pueda dedicar su tiempo a desarrollar tareas de supervisión y mayor generación de valor en su día a día.

Los sistemas de IA deberían respaldar la autonomía y la toma de decisiones de las personas, tal como prescribe el principio del respeto de la autonomía humana. Esto requiere que los sistemas de IA actúen tanto como facilitadores de una sociedad democrática, próspera y equitativa, apoyando la acción humana y promoviendo los derechos fundamentales, además de permitir la supervisión humana.

Los sistemas de IA pueden ser beneficiosos para las personas, por ejemplo, ayudándolas a llevar a cabo un seguimiento de sus datos personales o mejorando la accesibilidad de la educación, facilitando así el ejercicio del derecho a la educación. Sin embargo, dado el alcance y la capacidad de los sistemas de IA, también pueden afectar negativamente. En situaciones en las que existan riesgos de este tipo, deberá llevarse a cabo una evaluación del impacto. Además, deberían crearse mecanismos que permitan conocer las opiniones externas sobre los sistemas de IA que pueden vulnerar los derechos fundamentales.

Por otra parte, en ocasiones se pueden desplegar sistemas de IA con el objetivo de condicionar e influir en el comportamiento humano a través de mecanismos que pueden ser difíciles de detectar, explotando procesos del subconsciente mediante diversas formas de manipulación injusta, engaño, dirección y condicionamiento. Por todo esto, el principio general de autonomía del usuario debe ocupar un lugar central en la funcionalidad del sistema. La clave para ello es el derecho a no ser sometido a una decisión basada exclusivamente en procesos automatizados, cuando tal decisión produzca efectos jurídicos sobre los usuarios o les afecte de forma significativa, como sucede en lo relativo al empleo o a la salud.

Cabe distinguir aquí cuatro conceptos: supervisión humana, participación

humana, control humano y mando humano.

La **supervisión humana** ayuda a garantizar que un sistema de IA no socave la autonomía humana o provoque otros efectos adversos. La supervisión se puede llevar a cabo a través de mecanismos de gobernanza, tales como los enfoques de participación humana, control o mando humanos.

Por su parte, la **participación humana** hace referencia a la capacidad de que intervengan seres humanos en todos los ciclos de decisión del sistema, algo que en muchos casos no es posible ni deseable.

El **control humano** se refiere a la capacidad de que intervengan seres humanos durante el ciclo de diseño del sistema y en el seguimiento de su funcionamiento. Por último, el **mando humano** es la capacidad de supervisar la actividad global del sistema de IA (incluidos, desde un punto de vista más amplio, sus efectos económicos, sociales, jurídicos y éticos), así como la capacidad de decidir cómo y cuándo utilizar el sistema en una situación determinada. Esto puede incluir la decisión de no utilizar un sistema de IA en una situación particular, establecer niveles de discrecionalidad humana durante el uso del sistema o garantizar la posibilidad de ignorar una decisión adoptada por un sistema.

Además, se debe garantizar que los responsables públicos puedan ejercer la supervisión en consonancia con sus respectivos mandatos. Puede ser necesario introducir mecanismos de supervisión en diferentes grados para respaldar otras medidas de seguridad y control, dependiendo del ámbito de aplicación y el riesgo potencial del sistema de IA. Si el resto de las circunstancias no cambian, **cuanto menor sea el nivel de supervisión que pueda ejercer una persona sobre un sistema de IA, mayores y más exigentes serán las verificaciones y la gobernanza necesarias**.

Fuente: IBM Watson Health

En resumen, **la IA viene a mejorar el mundo en el que vivimos, no a que deleguemos en ella nuestra responsabilidad o a eliminar el factor humano de la ecuación**. Ese binomio persona-máquina será la clave en la evolución hacia una medicina de precisión, frente a un modelo de decisiones algorítmicas basadas en la elaboración de perfiles de pacientes. A través de la medicina de precisión, se podrá prevenir la asunción de conclusiones equivocadas o sesgadas, en las que también pueden incurrir los sistemas computacionales.

DESAFÍOS

1.- Asegurar que los sistemas de IA incorporan mecanismos para que los profesionales sanitarios puedan intervenir o supervisar cualquier punto del proceso de decisión, así como corregir cualquier decisión de dichos sistemas.

2.- Una supervisión adecuada de sistemas automatizados requiere un seguimiento continuo más allá de la certificación.

3.- Esa constante actividad de revisión y validación puede frustrar algunas de las expectativas depositadas en la IA para el ejercicio de la medicina.

4.- Asignación de los distintos niveles de responsabilidad (*accountability*) de todas las personas (también jurídicas) implicadas en el proceso, además del facultativo.

PROPUESTAS

1.- Es responsabilidad de todas las partes interesadas asegurarse de que los sistemas de IA se utilicen en las condiciones adecuadas y por personas debidamente capacitadas.

2.- Deben existir mecanismos efectivos de control, medición de impacto y reparación para las personas y grupos que se vean afectados negativamente por decisiones basadas en algoritmos.

3.- Particularmente en el ámbito sanitario, las personas deben mantener el control de los sistemas de atención de la salud y las decisiones médicas, de modo que la atribución de responsabilidades en caso de error de juicio del algoritmo nunca recaiga en un ente no humano.

4.- Para que la exigencia de revisiones y constantes controles no disuada a los profesionales sanitarios del uso de estas herramientas, conviene asimismo sensibilizar a éstos sobre la importancia de contribuir a la mejora progresiva de los mismos y la minimización del concepto de "riesgo aceptable", a través, precisamente, de la experiencia compartida de toda la profesión.

5.- Allá donde sea posible, por el bajo riesgo, elaborar protocolos de actuación (científicamente respaldados) para aligerar la carga de los facultativos y/o del personal administrativo.

6.- Complementariamente a la ya mencionada supervisión constante de los facultativos, o a un potencial esquema de certificaciones, puede ser pertinente contar con el apoyo de técnicos expertos acreditados e independientes que pudieran, por ejemplo, validar la adecuación del manejo de ciertas herramientas para cada uno de los usos para los cuales fue diseñada o contratada, y a los cuales pudieran acudir los médicos en caso de dudas a este respecto.

Enrique Sahún

Data & AI Manager en Nae

La necesaria transición hacia una Inteligencia Aumentada o Colaborativa

Uno de los mayores temores de la sociedad tiene ante el uso de la IA en el ámbito de la Salud es que alguna especie de máquina o robot acabe sustituyendo a los profesionales sanitarios, especialmente a nuestro médico o médica de confianza, de manera que perdamos ese trato cercano y humano que tenemos ahora. Además, debemos reconocer la desconfianza que todavía nos genera que un sistema artificial sea el único responsable de realizarnos un diagnóstico médico y prescribirnos un tratamiento de cualquier tipo.

Estas preocupaciones son perfectamente entendibles por el desconocimiento que existe todavía en gran parte de la sociedad acerca de la IA, lo cual sin duda se irá solucionando con el paso del tiempo, que ayudará a que todos veamos los beneficios que nos aportan estas tecnologías y podamos confiar más en ellas.

En el sentido más puramente pedagógico, **será fundamental hacer ver a la sociedad los esfuerzos que todo tipo de organizaciones públicas, privadas y civiles están haciendo para garantizar que los sistemas de IA cumplen con los principios éticos**, que a su vez están derivados de los derechos fundamentales de las personas. En el caso de la Unión Europea, ya se está trabajando en garantizar el cumplimiento de principios como el de prevención del daño, para evitar que un sistema de IA pueda provocar o agravar un daño en la salud de las personas; o el principio de explicabilidad, para asegurar la transparencia en la toma de decisiones de estos sistemas de IA, como puede ser un determinado diagnóstico o tratamiento. Pero también el foco está en garantizar el principio del respecto de la autonomía humana, que es, en mi opinión, en el que se deben poner los mayores esfuerzos en asegurar su cumplimiento de manera prioritaria.

Este principio es el que debe garantizar que las personas que interactúen con sistemas de IA (en este caso, los profesionales sanitarios) deben tener una autonomía plena y efectiva sobre sí mismas, al mismo tiempo que **los sistemas de IA no deberían subordinar, manipular o dirigir a estos profesionales de manera injustificada**. Para asegurar su cumplimiento, ya se trabaja en garantizar que los sistemas de IA tengan mecanismos de control y supervisión por parte de los profesionales sanitarios. Sin embargo, en mi opinión, es fundamental comprender que **los mayores beneficios los obtendremos siempre de la combinación de las capacidades de los sistemas de IA y de las personas, cada cual en lo que es realmente diferencial**. Los sistemas de IA, en la velocidad de análisis de grandes volúmenes de información, en el descubrimiento de patrones, en la identificación de anomalías, etc. Las personas, en el aspecto emocional, en el entendimiento del contexto y, especialmente, en la toma de la decisión final. Será así como pasaremos de hablar de IA a modelos más humanizados, pero también más potentes, como **Inteligencia Aumentada o Inteligencia Colaborativa**.

No me cabe duda de que en los próximos años vamos a ir viendo cómo la IA se aplica de manera más y más frecuente en el ámbito sanitario, pero siempre como una herramienta más que permitirá a los profesionales sanitarios hacer su trabajo de una manera más eficiente y precisa. Y seremos nosotros, los ciudadanos, los mayores beneficiados.


digitales_

6.- RESPONSABILIDAD & REGULACIÓN

Eva A. Kaili

Chair of STOA and the Centre for AI (C4AI), Parlamento Europeo

P.- ¿Cree que regular la IA podría resultar coercitivo para la innovación?

A nivel europeo, estamos trabajando en una legislación para regular los usos de alto riesgo de la IA, no la tecnología en sí. Esta legislación establecerá unos principios mínimos de cumplimiento y responsabilidad para garantizar que las aplicaciones de IA no generen ningún daño, exclusión o discriminación social. Este enfoque de "ética desde el diseño" es perfectamente compatible con la calidad del servicio al usuario y la innovación basada en el respeto de los derechos humanos fundamentales.

P.- ¿Qué rol puede jugar Europa en el establecimiento de esos estándares éticos básicos?

R.- Europa puede marcar el camino del resto, del mundo, igual que lo ha hecho en el ámbito de la protección de datos. Estamos trabajando con la OCDE para lograr el máximo consenso internacional posible en este ámbito.

Como hemos visto a lo largo de este informe, la cuestión de la atribución de responsabilidades preocupa e influye de forma notoria en la utilización de herramientas de IA para fines predictivos o diagnósticos en el campo médico. Ha quedado reflejado, asimismo, que las soluciones de IA implican nuevas fuentes de riesgos que pueden surgir en cualquiera de las etapas del ciclo de vida de las mismas (diseño, datos, entrenamiento, consumo). Por ello, se impera la necesidad de establecer normas, procedimientos y criterios, asignando responsabilidades y definiendo un marco de control para mitigar los riesgos de la IA de manera trasversal en la organización y a lo largo de toda la cadena de valor.

Por ello, los requisitos anteriores (privacidad, robustez, no discriminación, etc.) se complementan con el de rendición de cuentas, estrechamente relacionado con el principio de equidad. Este requisito exige establecer mecanismos que permitan garantizar la responsabilidad y *accountability* sobre los sistemas de IA y sus resultados, tanto antes de su implantación como después de ésta.

La **auditabilidad** es la capacidad para evaluar los algoritmos, los datos y los procesos de diseño. Esto no implica necesariamente que siempre deba disponerse de forma inmediata de la información sobre los modelos de negocio y la propiedad intelectual del sistema de IA. La evaluación por parte de auditores internos y externos y la disponibilidad de los correspondientes informes de evaluación pueden contribuir a la fiabilidad de esta tecnología. En aplicaciones que afecten a los derechos fundamentales, incluidas las aplicaciones esenciales desde el punto de vista de la seguridad, los sistemas de IA deberían poder someterse a auditorías independientes.

Es preciso garantizar tanto la capacidad de informar sobre las acciones o decisiones que contribuyen a un determinado resultado del sistema, como de responder a las consecuencias de dicho resultado. La identificación, evaluación, notificación y minimización de los posibles efectos negativos de los sistemas de IA resulta especialmente crucial para quienes resulten (in)directamente afectados por ellos. Debe protegerse debidamente a los denunciantes anónimos, las ONG, los sindicatos u otras entidades que trasladen preocupaciones legítimas en relación con un sistema basado en IA. La utilización de **evaluaciones de impacto** (como, por ejemplo, los «equipos rojos» o determinados tipos de evaluación algorítmica de impacto), antes y después del desarrollo, despliegue y utilización de sistemas de IA, puede resultar útil para minimizar sus efectos negativos. Estas evaluaciones deben ser proporcionadas al riesgo que planteen los sistemas.

A la hora de aplicar los requisitos anteriores, pueden surgir tensiones entre ellos, por lo que puede ser necesario buscar el equilibrio. Este tipo de situaciones deberían abordarse de manera racional y metódica de acuerdo con el nivel técnico actual. Esto significa que se deberían identificar los intereses y valores subyacentes al sistema de IA y que, **en el caso de que surjan conflictos, se deberá explicitar cómo se ha intentado buscar el equilibrio entre ellos y evaluar dicho equilibrio en términos del riesgo** que plantea para los principios éticos, incluidos los derechos fundamentales.

Llevar todo esto a la práctica no siempre resulta sencillo. Pueden darse situaciones en que no sea posible identificar equilibrios aceptables desde el punto de vista ético. En esos casos, no se debería continuar con el desarrollo, despliegue y utilización del sistema de IA en la forma prevista. Cualquier decisión sobre la búsqueda de equilibrios debe razonarse y documentarse convenientemente. El encargado de la adopción de decisiones debe ser responsable de la forma en que se busque el equilibrio en cuestión, y revisar constantemente la idoneidad de la decisión resultante para garantizar que se puedan introducir los cambios necesarios en el sistema cuando sea preciso.

Los profesionales consultados para la elaboración de este informe precisan, asimismo, que se producen algunas reticencias para el uso de IA por parte de los profesionales sanitarios causadas por la incertidumbre con respecto a su grado de responsabilidad en las decisiones adoptadas en base a un algoritmo.

Desde el punto de vista del paciente, cuando se produzcan efectos adversos injustos, deberían preverse mecanismos accesibles que aseguren una compensación adecuada. El hecho de saber que se podrá obtener una reparación si las cosas no salen según lo previsto es crucial para garantizar la confianza de la sociedad. Se debería prestar atención a las personas o grupos vulnerables, por razones de justicia social, y para no disuadir a estas personas de compartir sus datos, contribuyendo así a mejorar la robustez de los algoritmos y a reducir el riesgo de sesgos en las muestras.

Por descontado, la rendición de cuentas es una labor muy *humana* y la asignación de las distintas responsabilidades ha de recaer siempre en personas físicas o jurídicas.

La reflexión ética vive en un estado previo, paralelo y complementario a la regulación. Esta última resulta efectiva para establecer y **homogeneizar los mecanismos mínimos de control, supervisión, trazabilidad y auditabilidad**. Dada la rapidez con la que avanza el desarrollo tecnológico, cualquier texto regulatorio debe procurar un equilibrio entre la flexibilidad que necesita una norma en un mundo tan cambiante como el actual, y la concreción en el articulado allá donde pueden establecerse criterios técnicos, ya que estos ayudan a homogeneizar las obligaciones y procedimientos.

En términos generales, la IA no puede ser ajena a los altos estándares que el sistema sanitario aplica sobre cualquier otra tecnología médica. Así, el personal sanitario recibe una exhaustiva formación continuada y sus decisiones y conocimientos son permanentemente fiscalizados, mientras que los productos y medicamentos que dispensan son probados y validados en procesos igualmente exigentes.

Algunas de las preguntas abiertas al debate sobre la responsabilidad y los límites de la regulación son: ¿Y si un nivel más alto de transparencia no ayudara a crear sistemas de IA mejor interpretables por las personas? ¿Y si detrás de las propuestas y regulaciones en torno al

derecho a la información y a la explicación escondiéramos suposiciones (sesgos) implícitas? **¿Necesitan estas herramientas de un grado de transparencia superior al que exigimos a los responsables de las tomas de decisiones humanas?**

Por último, cuando la elaboración de perfiles o las decisiones automatizadas afectan a la capacidad de las personas para acceder a determinados medicamentos, tratamientos, etc., se debate dónde establecer la frontera entre el derecho a la información y el derecho a la explicación. El primero está ampliamente garantizado en la legislación española, si bien la decisión de hasta dónde llega el límite máximo del riesgo permitido continúa a discrecionalidad de los responsables de cada proyecto. Por lo demás, la regulación ya existente es exigente y eficaz.

DESAFÍOS

1.- Hay situaciones donde la excesiva regulación bloquea el desarrollo y crecimiento de la Inteligencia Artificial. Se debe conseguir que regulación y desarrollo crezcan y evolucionen paralelamente de forma eficiente para buscar el máximo progreso posible.

2.- Existencia extendida de un Comité de Ética en las empresas [sanitarias], que produzcan usos y soluciones de IA en sanidad.

3.- Definir inequívocamente a quién corresponde la responsabilidad y rendición de cuentas ante un daño provocado por estos sistemas.

4.- Establecer quién determina el nivel de riesgo que se atribuye al sistema inteligente, antes de su introducción en el mercado, y qué margen de discrecionalidad le queda al facultativo para aceptar o rechazar ese punto de partida.

5.- Es posible que algunas tecnologías de IA no emitan una sola decisión, sino un conjunto de opciones entre las que un médico debe seleccionar. En ocasiones, el desarrollador, la institución y el médico pueden haber jugado un papel en el daño médico, de modo que ninguno puede ser culpado.

6.- Cabe reflexionar si el paciente debiese ser conocedor del uso de IA antes de proporcionar un "consentimiento informado", y si la ausencia de esa información podría implicar una responsabilidad legal para el profesional, en caso de un resultado insatisfactorio. Este proceso de transparencia resulta particularmente desafiante en los consentimientos electrónicos para servicios digitales, donde la interacción humana es menor.

7.- Los algoritmos complejos de aprendizaje automático pueden resultar difíciles de entender para los reguladores.

PROPUESTAS

1.- Fomento de propuestas de autocontrol.

2.- Promover una regulación con enfoque basado en el riesgo aceptable para fomentar la confianza en la IA sin obstaculizar su desarrollo responsable. DigitalES aboga por un enfoque proporcionado, que regule los casos de uso de alto riesgo, y no la tecnología de Inteligencia Artificial en sí.

3.- Las aplicaciones de IA deben considerarse de alto riesgo cuando cumplen con ciertos criterios, como la probabilidad de que ocurran daños graves a las personas (por ejemplo, amenazas a la salud, la vida o los derechos fundamentales).

4.- Trabajar en un marco que defina sin ambigüedades los límites de responsabilidad de los desarrolladores de sistemas de IA para el ámbito sanitario, y que establezca mecanismos de traspaso de la responsabilidad a los profesionales sanitarios en casos donde deban ser estos profesionales quienes prevalezcan sobre los sistemas de IA, por el potencial daño que podría ocasionar a las personas un mal desempeño de estos últimos.

5.- Promover que las organizaciones utilicen normas armonizadas y la autoevaluación de la conformidad de algunos de sus productos.

REGULACIÓN EUROPEA DE LA IA - ¿EN QUÉ FASE SE ENCUENTRA?

En mayo, la Unión Europea se convirtió en el primer organismo gubernamental del mundo en emitir una respuesta integral en forma de borradores de reglamentos destinados específicamente al desarrollo y uso de la IA. Las regulaciones propuestas se aplicarían a cualquier sistema de IA utilizado o que proporcione resultados dentro de la Unión Europea, lo que indica implicaciones para las organizaciones de todo el mundo. Aunque la regulación de la UE aún no está en vigor, proporciona una visión clara del futuro de la regulación de la IA en su conjunto.

Desde que la Comisión Europea presentó su proyecto de Ley de IA, el progreso legislativo ha sido lento. En parte porque el archivo es muy técnico, los responsables de la formulación de políticas han tenido conflictos para comprender las implicaciones de las disposiciones legales. La regulación de la IA también interactúa con otras leyes de la UE, desde la protección de datos hasta la seguridad de los productos a través de la aplicación de la ley. Esas, entre otras razones hizo que la anterior presidencia eslovena del Consejo proporcionase solo una reescritura limitada de la propuesta, tratando de darle forma en algunos aspectos críticos antes de pasar el dossier a la nueva presidencia francesa. En el seno del Parlamento Europeo, la decisión sobre las comisiones con competencia legislativa en el dossier ha hecho también que los tiempos se alarguen.

Respecto a su entrada en vigor, aunque no hay forma de saberlo con seguridad, el cronograma para la adopción de GDPR, que se propuso en 2012, se adoptó en 2014 y entró en vigor en 2018, podría brindar alguna orientación. Sin embargo, independientemente de la línea de tiempo, existen muchas leyes y regulaciones que ya se aplican al uso de la IA, especialmente porque algunas de ellas son específicas del sector y no hacen referencia explícita a la IA. En la Unión Europea, por ejemplo, GDPR ya requiere el consentimiento explícito de las personas antes de que estén sujetas a decisiones basadas únicamente en el procesamiento automatizado. Así el proyecto de reglamento de la UE es un paso más en lo que se convertirá en un esfuerzo global para gestionar los riesgos asociados con la IA.

Pablo Fernández Burgueño

Abogado y senior manager del equipo NewLaw en PwC Tax & Legal

Detrás de un sistema de IA siempre hay personas

Las soluciones de Inteligencia Artificial (IA) no son, en sí mismas, responsables desde el punto de vista ético, pero su mera existencia y su funcionamiento sí generan responsabilidad en las personas con la consecuente rendición de cuentas.

La búsqueda de soluciones de IA se desarrolla en un joven marco de innovación a través del cual se trata de encontrar sistemas informáticos que presenten las mismas capacidades que los seres humanos, aunque se quiere ir más allá. Bajo este paraguas, el ser humano ha logrado programar sistemas capaces de replicar situaciones, predecir eventos y lograr la abstracción suficiente para resolver problemas de extraordinaria complejidad. Los logros irrefutables y, muchas veces, inexplicables marcan los siguientes pasos creativos en este campo de investigación.

La acción de la IA sobre diferentes campos y, en particular, sobre el sector de la salud genera impactos de diferente entidad en todas las facetas de la vida, el medioambiente, el respeto a los derechos humanos, el buen gobierno y, en definitiva, sobre el futuro de la humanidad. Cada impulso eléctrico de la IA, sea leve o fuerte (sea un 0 o un 1) o sea cuántico, representa el batir de una mariposa que aún el ser humano no es capaz de comprender y que cambia el rumbo y destino de los individuos y la sociedad. **Desde la fase de ideación, en la que la IA es diseñada por personas -o por otras IA diseñadas por personas-, hay seres humanos, empresas y gobiernos sobre los que pesa el deber** de crear y usar sobre sólidos pilares éticos, de cumplir con la normativa en vigor y de asumir la responsabilidad por sus actos y las consecuencias de estos.

En el proceso creativo, con sus correspondientes fases de ideación, desarrollo, prueba, funcionamiento e iteración de la IA, participa un gran número multidisciplinar de profesionales y todos ellos asumen la debida responsabilidad sobre su aportación. Médicos, abogados, informáticos y humanistas, entre otros, aportan conocimiento en cada una de sus acciones. El objetivo a veces es crear una IA que dé solución a un problema, y otras es eficientar un sistema, o implementar una nueva funcionalidad. Cada persona se une a este trabajo creativo aportando según quien es y lo que profesionalmente ejerce, con sus diferentes intereses, bases culturales y grado personal de implicación. Tan posible es que una IA genere resultados faltos de ética por una mala decisión tomada en la fase de ideación por una persona implicada en el diseño, que por un usuario en la fase en la que la solución ya está en funcionamiento.

Es necesario evitar el riesgo, pero también lo es identificar su origen, ya sea para anularlo o para mitigarlo. Cuestionar la propia idea del riesgo es posible gracias al despliegue de medidas preventivas tales como la elaboración de auditorías y evaluaciones de impacto. Sin embargo, en materia de IA, conocer el origen del riesgo a veces no permitirá atacarlo, sino comprenderlo e identificar al responsable causal.

Toda acción humana genera consecuencias jurídicas, y la IA, cuya existencia se debe a la acción humana, no escapa de esta máxima. **Una autorregulación responsable en el marco de la IA ética y el cumplimiento de la normativa**, junto con una necesaria labor de **creación de prueba documental sobre cada uno de los eventos producidos para y por la IA**, permitirán construir y afianzar un marco de rendición de cuentas éticamente responsable.


digitalgs_

RESUMEN DE RECOMENDACIONES

TRANSPARENCIA & EXPLICABILIDAD

1.- Promover obligaciones mínimas de transparencia para aplicaciones de menor riesgo: las personas tienen derecho a saber cuándo están interactuando con un sistema de IA.

2.- Definir criterios mínimos de explicabilidad que deban ofrecer los sistemas de IA a los profesionales sanitarios. De forma más concreta, es importante asegurar la trazabilidad sobre:

- Los métodos utilizados para diseñar y desarrollar el sistema algorítmico.
- Los métodos empleados para ensayar y validar el sistema algorítmico.
- Los resultados del sistema algorítmico.

Garantizar la trazabilidad significa que las decisiones ejecutadas por sistemas algorítmicos puedan ser auditadas, evaluadas y explicadas por las personas responsables.

3.- Proporcionar desde las instituciones formación sobre el funcionamiento de la IA en abierto.

4.- Incorporar en el grado de Medicina asignaturas sobre salud digital para actualizar el grado a las demandas contemporáneas de la medicina.

5.- Establecimiento de una responsabilidad legal ante la no transparencia en los sistemas.

PRIVACIDAD

6.- Otorgar la formación y conocimientos necesarios a las empresas/centros sanitarios para asegurar el cumplimiento de la normativa en protección datos, evitando confusiones o actuaciones ilícitas fuera del margen legal.

7.- Crear mecanismos que aseguren que el desconocimiento de la ley no desemboca en una actividad ilegal.

8.- Incorporar mecanismos robustos de pseudo-anonimización de los datos que garanticen la privacidad de éstos a la vez que pueden ser utilizados para fines de investigación de nuevos medicamentos, tratamiento de enfermedades, etcétera.

9.- Desarrollar planes de contingencia en caso de filtración, robo, pérdida o re-identificación de los datos.

CALIDAD DEL DATO & DISCRIMINACIÓN

10.- Promover iniciativas de formación y capacitación a toda la población, especialmente la más vulnerable, sobre el uso de herramientas basadas en IA para el ámbito sanitario.

11.- Requerir a los desarrolladores de sistemas de IA la inclusión de mecanismos de control, incluyendo la intervención humana, que garanticen la no presencia de sesgo de ningún tipo en sus soluciones.

12.- Acelerar la interoperabilidad de sistemas, asegurando la privacidad y la calidad de los datos, para ampliar el tamaño de la muestra de información que vaya a ser procesada. En esta línea, España está trabajando ya en la creación de un data lake (repositorio único) de datos sanitarios.

13.- Incorporar a las acciones de formación y explicabilidad dirigidas a personal sanitario advertencias sobre los posibles sesgos de la información que manejan.

ROBUSTEZ

14.- Definir unos niveles mínimos de fiabilidad y precisión para cada tipo de aplicación de los sistemas de IA que se deban cumplir antes de ponerlos en producción, en función de la sensibilidad del caso de uso.

15.- Promover la implantación de un sistema de protección o de almacenamiento (nube) que permita crear copias de seguridad y proteger los datos de pacientes ante cualquier tipo de inconveniente, ya sea apagón, error o destrucción del sistema, ataques de seguridad.

16.- Promover un sistema de certificación para software de IA en sanidad a nivel nacional, basado en criterios objetivos de calidad de las auditorías algorítmicas, e impulsar esa certificación a nivel europeo.

17.- No todos los modelos algorítmicos ofrecen los mismos resultados en los distintos casos de uso. El expertise en el ámbito health (y la experiencia previa de otras investigaciones y organizaciones) será esencial para saber escoger entre Machine Learning y Deep Learning y, dentro de la primera categoría, entre aprendizaje supervisado, aprendizaje no supervisado y aprendizaje reforzado.

AUTONOMÍA & SUPERVISIÓN HUMANA

18.- Es responsabilidad de todas las partes interesadas asegurarse de que los sistemas de IA se utilicen en las condiciones adecuadas y por personas debidamente capacitadas.

19.- Deben existir mecanismos efectivos de control, medición de impacto y reparación para las personas y grupos que se vean afectados negativamente por decisiones basadas en algoritmos.

20.- Particularmente en el ámbito sanitario, las personas deben mantener el control de los sistemas de atención de la salud y las decisiones médicas, de modo

que la atribución de responsabilidades en caso de error de juicio del algoritmo nunca recaiga en un ente no humano.

21.- Para que la exigencia de revisiones y constantes controles no disuada a los profesionales sanitarios del uso de estas herramientas, conviene asimismo sensibilizar a éstos sobre la importancia de contribuir a la mejora progresiva de los mismos y la minimización del concepto de "riesgo aceptable", a través, precisamente, de la experiencia compartida de toda la profesión.

22.- Allá donde sea posible, por el bajo riesgo, elaborar protocolos de actuación (científicamente respaldados) para aligerar la carga de los facultativos y/o del personal administrativo.

23.- Complementariamente a la ya mencionada supervisión constante de los facultativos, o a un potencial esquema de certificaciones, puede ser pertinente contar con el apoyo de técnicos expertos acreditados e independientes que pudieran, por ejemplo, validar la adecuación del manejo de ciertas herramientas para cada uno de los usos para los cuales fue diseñada o contratada, y a los cuales pudieran acudir los médicos en caso de dudas a este respecto.

REGULACIÓN & RESPONSABILIDAD

24.- Fomento de propuestas de autocontrol.

25.- Promover una regulación con enfoque basado en el riesgo necesario para fomentar la confianza en la IA sin obstaculizar su desarrollo responsable. DigitalES ha solicitado tradicionalmente un enfoque proporcionado, que regule los casos de uso de alto riesgo, y no la tecnología de Inteligencia Artificial en sí.

26.- Promover que las organizaciones utilicen normas armonizadas y la autoevaluación de la conformidad de algunos de sus productos.

27.- Las aplicaciones de IA deben considerarse de alto riesgo cuando cumplen con ciertos criterios, como la gravedad y la probabilidad de que ocurran ciertos daños graves a las personas (por ejemplo, amenazas a la salud, la vida o los derechos fundamentales).

28.- Se debe trabajar en un marco que defina sin ambigüedades los límites de responsabilidad que tendrían los desarrolladores de sistemas de IA para el ámbito sanitario. También, que establezca mecanismos de traspaso de la responsabilidad a los profesionales sanitarios en casos donde deban ser estos profesionales quienes prevalezcan sobre los sistemas de IA, por el potencial daño que podría ocasionar a las personas un mal desempeño de estos últimos.

CONCLUSIONES

El profesional sanitario debe ser parte íntegra y esencial de los sistemas de Inteligencia Artificial. Debe estar involucrado desde la comprensión de los resultados que genera la IA.

Se debe asegurar el cumplimiento de los principios éticos para conseguir una Inteligencia Artificial de calidad que sea respetuosa con el paciente y alcance los objetivos establecidos.

La Inteligencia Artificial debe ser concebida por el médico como una herramienta más a su disposición y apoyarse en los resultados que genera para tomar decisiones.

Es necesaria la formación de los profesionales sanitarios para una correcta interacción con los sistemas de Inteligencia Artificial.

La Inteligencia Artificial debe generar confianza en la población, que entienda sus beneficios y anime al paciente a colaborar en la mejora de su desarrollo a través de la facilitación de sus datos de salud.

La ética y el desarrollo de IA deben ir de la mano: la ética no debe frenar el desarrollo tecnológico de la IA. Al igual que el desarrollo de la IA no puede ir por delante del cumplimiento ético y atentar contra la privacidad o derechos de los ciudadanos.

ANEXO

BIENESTAR SOCIAL & FUTURO DEL EMPLEO

Melvin Kranzberg, que fue profesor de historia de la tecnología en el Instituto de Tecnología de Georgia y un reconocido teórico sobre la materia, sostenía que *"la tecnología no es buena ni mala, ni es neutra*[5]*"*. Kranzberg nos recordó también que *"la tecnología es una actividad muy humana"*. Un siglo antes, Víctor Hugo dio con la clave:

"El futuro tiene muchos nombres. Para los débiles es lo inalcanzable. Para los temerosos, lo desconocido. Para los valientes es la oportunidad"

La ética en el desarrollo tecnológico está directamente ligada con la confianza y, en consecuencia, con el uso real de estas soluciones. Está ligada también con la inclusividad y la sostenibilidad de los países, en un mundo crecientemente conectado y globalizado.

Diversos estudios posicionan a la IA como la tecnología con mayor potencial para fomentar el progreso social y económico. A estas oportunidades se suman una serie de retos éticos al abordar las implicaciones sociales del uso inadecuado de la tecnología. No atender a ciertos riesgos de la IA afectar los derechos fundamentales, restringir el acceso a

oportunidades y ralentizar el desarrollo social.

En consonancia con los principios de equidad y prevención del daño, se debería tener en cuenta también a la sociedad en su conjunto y al medio ambiente como partes interesadas a lo largo de todo el ciclo de vida de la IA. **Un uso responsable de la IA debe por un lado fomentar la sostenibilidad y la responsabilidad ecológica de los sistemas de IA, y también abarcar nuevos ámbitos con impacto social**, impulsando la investigación de soluciones de inteligencia artificial para hacer frente a los temas que suscitan preocupación a escala mundial, como los Objetivos de Desarrollo Sostenible. Esta visión persigue que la IA se utilice en beneficio de todos los seres humanos, incluidas las generaciones futuras.

Los sistemas de inteligencia artificial prometen ayudar a abordar algunas de las preocupaciones sociales más urgentes. No obstante, se debe garantizar que lo hagan del modo más respetuoso posible con el medio ambiente y siendo sostenibles. En ese sentido, debería evaluarse en su integridad el proceso de desarrollo, despliegue y utilización de sistemas de IA, así como toda su cadena de suministro, a través, por ejemplo, de un examen crítico del uso de los recursos y del consumo de energía a lo largo de la cadena de suministro, dando prioridad a las opciones menos perjudiciales. Se deberían promover medidas que garanticen el respeto del medio ambiente por parte de todos los eslabones de la cadena de suministro.

Las organizaciones que desarrollan aplicaciones de Inteligencia Artificial deben cuidar y ser consciente de que éstas generan un impacto social. La exposición ubicua a los sistemas sociales de IA en todas las esferas de nuestra vida (sea en ámbitos como la educación, el trabajo, el entretenimiento, el cuidado o la sanidad) pueden alterar nuestra concepción de la acción social o afectar a nuestras relaciones y vínculos sociales. Aunque los sistemas de IA se pueden utilizar para mejorar las competencias sociales, también pueden contribuir a su deterioro. Por lo tanto, será necesario tener en cuenta y llevar a cabo un seguimiento exhaustivo de los efectos de esos sistemas.

En este contexto, la Estrategia Nacional de Inteligencia Artificial[6], presentada en diciembre de 2020, propone "evaluar si nuestras normas de convivencia están adaptadas a las necesidades del momento, si es suficiente con el marco ético y jurídico que nos ha acompañado hasta hoy o qué ajustes y revisiones necesita para preservar los derechos de la ciudadanía en un mundo digital y anteponer objetivos éticos y democráticos al desarrollo de la IA".

De su lado, la Organización Mundial de la Salud (OMS) propone[7] realizar una investigación global de los efectos sociales del uso de la IA en el ámbito sanitario. Algunas preguntas pertinentes sobre las que basar dicha investigación serían:

- ¿Para qué necesidades y deficiencias identificadas por los trabajadores de la

[6] https://portal.mineco.gob.es/es-es/ministerio/areas-prioritarias/Paginas/inteligencia-artificial.aspx

[7] https://www.who.int/news/item/28-06-2021-who-issues-first-global-report-on-ai-in-health-and-six-guiding-principles-for-its-design-and-use

salud y los pacientes podría la Inteligencia Artificial desempeñar un papel para garantizar la prestación de una atención equitativa?

- ¿Cómo está cambiando la Inteligencia Artificial las relaciones entre los profesionales sanitarios y los pacientes? ¿Estas tecnologías permiten a los proveedores dedicar más tiempo de "calidad" a los pacientes o hacen que la atención sea menos humana? ¿Los factores contextuales específicos mejoran o socavan la calidad de la atención?

- ¿Cuáles son las actitudes de los trabajadores sanitarios y los pacientes hacia el uso de la Inteligencia Artificial? ¿Consideran aceptables estas tecnologías? ¿Dependen sus actitudes del tipo de intervención, la ubicación de la intervención o la aceptación actual de estas tecnologías, tanto en el sistema sanitario como en la sociedad?

- ¿La introducción de la Inteligencia Artificial en el ámbito sanitario aumenta o reduce la brecha digital de la sociedad?

- ¿Cuál es la mejor manera de que los proveedores y programadores aborden los sesgos que se manifestarán en las aplicaciones? ¿Cuáles son las barreras para abordar los prejuicios?

- ¿Qué método debería utilizarse para evaluar si la Inteligencia Artificial es más rentable y apropiada que las soluciones preexistentes?

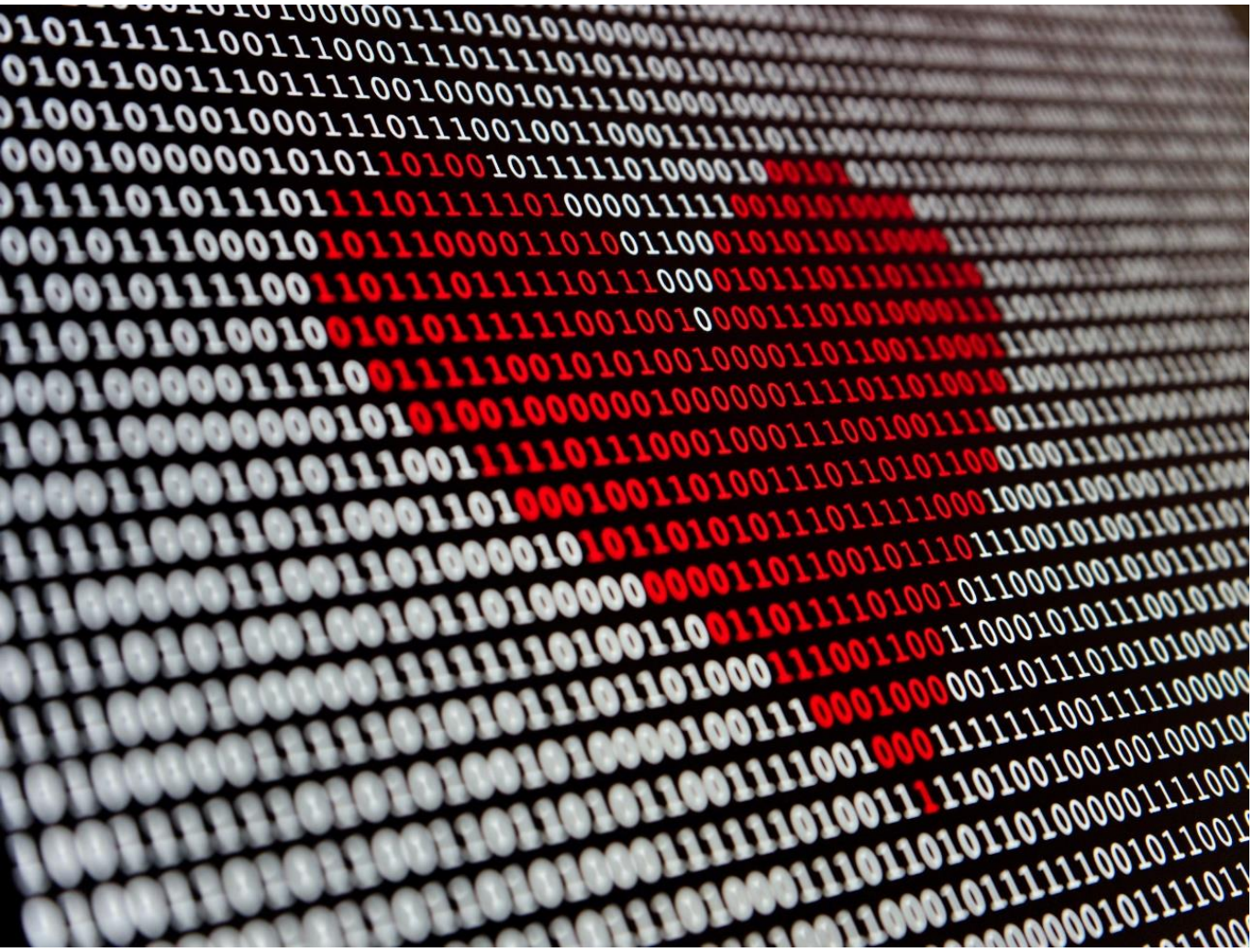
POSIBLE INCIDENCIA EN EL EMPLEO DEL SECTOR SALUD

Es un hecho que existe una preocupación y una expectativa crecientes por el futuro del trabajo en nuestras sociedades. A lo largo de los dos últimos siglos, el progreso técnico ha sido también motivo de previsiones lúgubres sobre la sustitución de los trabajadores por máquinas, la posibilidad de desempleo tecnológico y la participación del trabajo en la distribución de la renta. Sin embargo, aunque no sin tensiones ni conflictos, la historia de las sucesivas revoluciones industriales ha seguido hasta ahora un patrón de mejora generalizada del bienestar a largo plazo en las economías avanzadas.

¿Será la revolución digital diferente a las anteriores? El peor de los escenarios que algunos investigadores auguran es que el ritmo de avance tecnológico se debilite y derive en avances espectaculares (y aún hoy difícilmente predecibles) en tecnologías disruptivas y sesgadas en habilidades y tareas, que potencien la eficiencia de unos pocos trabajadores, mientras sustituyen y destruyen muchos otros empleos. Se produciría, por lo tanto, una "polarización" de la fuerza de trabajo y un "vaciado" de la clase media[8].

La mayoría de los estudios recientes, no obstante, matizan este escenario catastrófico de paro tecnológico masivo y desigualdad creciente, incidiendo en la importancia de que los profesionales

[8] https://www.technologyreview.es//s/3615/de-como-la-tecnologia-esta-destruyendo-el-empleo



potencien sus habilidades *soft* (el pensamiento crítico, la capacidad analítica, la creatividad, habilidades de trabajo en equipo y, en general, las competencias no cognitivas), que son precisamente aquellas de las que las máquinas adolecen.

Estar en posesión de las competencias requeridas en el futuro del trabajo será determinante de las probabilidades de inclusión o exclusión laboral de las personas, así como de la capacidad de los países de competir exitosamente en el mercado global.

Según un artículo el sistema británico de salud (NHS), en dos décadas, el 90% de todos los trabajos en el NHS requerirán habilidades digitales[9]. Además, el requisito de alfabetización digital se extenderá a los trabajadores en entornos determinantes para la salud, relacionadas con el entorno y el estilo de vida, como la vigilancia, el medio ambiente, la prevención, la nutrición, etcétera.

[9] The Topol review: Preparing the healthcare workforce to deliver the digital future. The Topol Review — NHS Health Education England (hee.nhs.uk)

Las opiniones más optimistas apuntan a que la IA automatizará y, por lo tanto, reducirá la carga de tareas rutinarias en los médicos, lo que les permitirá concentrarse en trabajos más desafiantes y en interactuar con los pacientes.

Los estudios que han tratado de cuantificar el efecto sustitución por causa de la automatización, debido al uso de diferentes metodologías, dan lugar a resultados muy diversos. Ahora bien, las investigaciones coinciden en que la probabilidad de automatización disminuye cuanto mayor es el grado de responsabilidad, el nivel educativo, la participación en actividades formativas, así como con la adopción de nuevas formas de trabajo.

Las investigaciones más recientes, como las que publica la OCDE[10], vienen así a poner el énfasis no tanto en la pérdida de empleo como en la tendencia hacia una transformación de las ocupaciones, donde se conjuguen tareas automatizadas, que se encarguen del trabajo más rutinario e incluso, en muchos casos, que suponga mayor riesgo para la seguridad y la salud en el trabajo, con el trabajo humano, que complemente la automatización con tareas menos rutinizables y acordes con los requerimientos de relaciones interpersonales, creatividad e innovación. En última instancia, el riesgo de automatización de los empleos no implica directamente la destrucción de los mismos.

Además, existe una alta probabilidad de que los avances tecnológicos, debido al "efecto compensación", terminen generando más crecimiento económico y empleo en el largo plazo.

Por otra parte, los aumentos de productividad que genera el uso de las nuevas tecnologías suelen llevar asociada una reducción de las horas de trabajo. Es, pues, difícil predecir los efectos de la automatización en el futuro del empleo en términos netos. En cualquier caso, las estimaciones realizadas hasta el momento difieren de manera importante, por lo que han de tomarse con cautela.

En definitiva, la IA se presenta como una **tecnología exponencial que maximiza la productividad y las capacidades de los seres humanos**. Se trata, por lo tanto, de una tecnología más encaminada a complementar esas capacidades que a sustituirlas, contribuyendo a alcanzar objetivos compartidos: mejora de la calidad asistencial, reducción de errores y la eficiencia del sistema sanitario.

DigitalES, Asociación Española para la Digitalización, reúne a las principales empresas del sector de la tecnología e innovación digital en España. El objetivo de DigitalES es impulsar la transformación digital, contribuyendo así al crecimiento económico y social de nuestro país. En conjunto, las empresas que forman parte de DigitalES facturan en España el equivalente a más del 3,4% del PIB nacional. Más información: www.digitales.es

[10] https://www.oecd.org/future-of-work/reports-and-data/

DIGITALES

Estamos presentes para crear el futuro

www.digitales.es