

Inteligencia artificial confiable y su uso en salud

Las decisiones tomadas por algoritmos en el diagnóstico o la planificación de tratamientos afectan directamente a la vida de las personas. Por ello es fundamental que los profesionales de la salud, y en última instancia los pacientes, comprendan cómo funcionan estos modelos y confíen en ellos, garantizando que los resultados de los algoritmos sean útiles y, sobre todo, seguros para los pacientes.



Pedro A. Moreno Sánchez

Doctor en Ingeniería Biomédica.
Senior Researcher en Tecnia Research & Innovation.

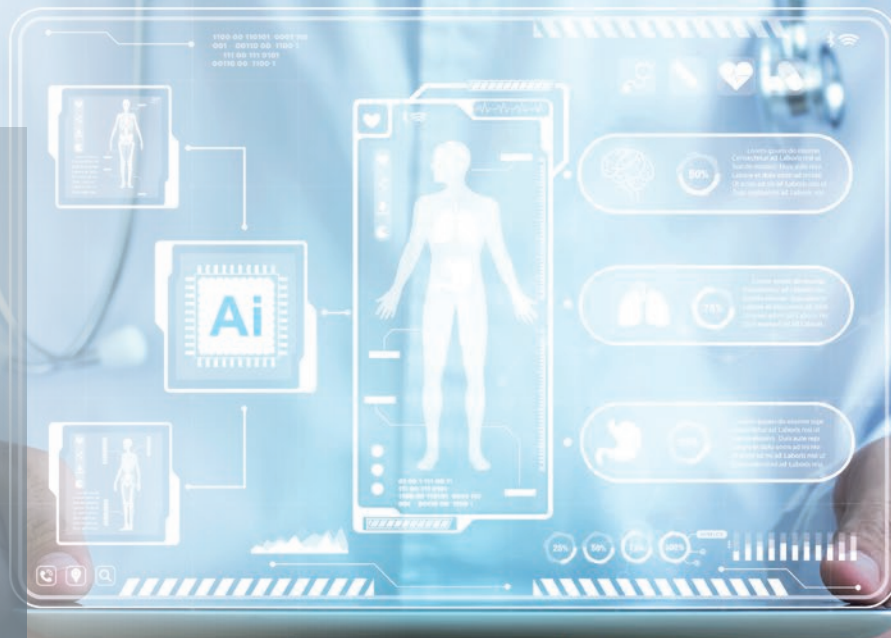
La inteligencia artificial (IA) está adquiriendo un papel cada vez más relevante en el ámbito de la salud, con aplicaciones que hace algunos años parecían impensables. En la actualidad, uno de los principales avances ha sido la implementación del aprendizaje profundo, una rama de la IA que permite analizar grandes volúmenes de datos para realizar tareas complejas como la detección de

enfermedades y la predicción de resultados clínicos. Estas técnicas permiten lograr diagnósticos más precisos y predicciones en ámbitos que antes resultaban inaccesibles para los métodos tradicionales. Un aspecto clave de la IA moderna es que aprende directamente de los datos, sin depender de reglas estrictas definidas por expertos. Esto permite descubrir relaciones ocultas en los resultados clínicos y explorar patrones no detectables de otro modo, facilitando la predicción de resultados de salud.

Sin embargo, uno de los principales desafíos es que los datos médicos, como las imágenes clínicas, no siempre están disponibles en la cantidad y calidad necesarias para entrenar modelos avanzados de aprendizaje automático. Los datos utilizados para entrenar los algoritmos de IA se basan en clasificaciones asignadas por el personal clínico, que en ocasiones puede contener imprecisiones o errores, lo que influye negativamente en las decisiones que toma el algoritmo. Además, la obtención de estos datos conlleva dificultades adicionales, como la necesidad de obtener el con-



“Los datos utilizados para entrenar los algoritmos de IA se basan en clasificaciones asignadas por el personal clínico, que en ocasiones pueden contener imprecisiones o errores, lo que influye negativamente en las decisiones que toma el algoritmo.”



sentimiento de los pacientes y superar diversas barreras administrativas. A medida que los profesionales del sector sanitario reconocen el valor de estos datos, los algoritmos y modelos de IA continúan mejorando, lo que permite diagnósticos más precisos y tratamientos personalizados.

La implementación de la IA en el ámbito de la salud plantea importantes retos, ya que las decisiones tomadas por algoritmos en el diagnóstico o la planificación de tratamientos afectan directamente a la vida de las personas. Por ello es fundamental que los profesionales de la salud,

y en última instancia los pacientes, comprendan cómo funcionan estos modelos y confíen en ellos, garantizando que los resultados de los algoritmos sean útiles y, sobre todo, seguros para los pacientes.

Por ejemplo, no es suficiente con obtener una predicción simple acerca del diagnóstico de una enfermedad; los médicos necesitan entender las razones que hay detrás de estas decisiones para poder evaluarlas críticamente. Además, existen preocupaciones clave relacionadas con la robustez de los modelos, la protección de la privacidad de los datos



“Se implementan mecanismos que permiten la intervención o anulación de una decisión tomada por un sistema de IA cuando sea necesario. Esto asegura que las decisiones automatizadas no reemplacen el juicio humano, sino que lo complementen de manera ética y responsable.”

y la equidad en las decisiones. Por último, surge la cuestión de la responsabilidad: si algo sale mal en una decisión automatizada, ¿quién debe rendir cuentas? Abordar estos aspectos es esencial para que la IA en salud se desarrolle de manera ética y confiable. Para abordar este desafío, la Unión Europea ha elaborado las “Guías éticas para la IA confiable”, un conjunto de directrices generalistas que buscan asegurar que el desarrollo y el uso de la IA sean responsables y confiables a través de 7 requisitos.

1. Derechos fundamentales, acción y supervisión humanas

El requisito de derechos fundamentales, acción y supervisión humanas en la IA en salud no solo garantiza que las decisiones tomadas por los algoritmos sean supervisadas por profesionales sanitarios, sino que también protejan los derechos fundamentales de los individuos, ya sean clínicos o pacientes. Es necesario evaluar aquellas situaciones donde las decisiones de IA puedan afectar a aspectos y derechos críticos, como es el derecho a la salud. Por otra parte, la IA debe empoderar a los profesionales y a los pacientes, apoyándolos para tomar decisiones autónomas e informadas, siempre de acuerdo con sus objetivos personales y clínicos. Un aspecto clave es el derecho a no ser

sometido a decisiones basadas únicamente en procesos automatizados cuando estas puedan tener efectos significativos en su vida. Para asegurar que la IA no socave la autonomía humana, se implementan mecanismos que permiten la intervención o anulación de una decisión tomada por un sistema de IA cuando sea necesario. Esto asegura que las decisiones automatizadas no reemplacen el juicio humano, sino que lo complementen de manera ética y responsable.

2. Robustez técnica y seguridad

La robustez técnica y seguridad en la IA aplicada a la salud es fundamental, dado que cualquier fallo en el sistema puede impactar directamente en la vida de los pacientes. Los sistemas de IA deben ser resistentes a ataques y manipulaciones, como la alteración de datos médicos, que puedan comprometer su funcionamiento. Además, es crucial que cuenten con planes de contingencia para mitigar riesgos en situaciones críticas. La precisión es otro aspecto crítico, especialmente cuando se trata de decisiones que pueden afectar a vidas humanas. Aunque es imposible evitar completamente predicciones inexactas, los sistemas de IA deben ser capaces de indicar la proba-

bilidad de error, lo que permite a los profesionales de la salud actuar con mayor cautela. Por último, la fiabilidad y la reproducibilidad son esenciales para asegurar que el sistema funcione correctamente en una variedad de casos y produzca resultados consistentes y confiables

3. Privacidad y gobernanza de los datos

Un factor crucial es la privacidad y gobernanza del dato en la IA aplicada a la salud, ya que se maneja información extremadamente sensible, como historiales médicos y datos personales de los pacientes. Es vital que estos sistemas aseguren la protección de la privacidad en todas las etapas del proceso, desde la recopilación hasta el almacenamiento y uso de los datos. Además, es necesario garantizar la calidad e integridad de los datos médicos, ya que cualquier error o sesgo puede afectar a los diagnósticos o decisiones clínicas. También deben existir protocolos estrictos que regulen quién puede acceder a estos datos y en qué circunstancias, para proteger la confidencialidad del paciente y asegurar un manejo ético y seguro de la información

4. Transparencia

La transparencia es uno de los requisitos más importantes para generar confianza y asegurar el uso correcto de la IA en salud. Los sistemas de IA deben ser comprensibles para los profesionales del sector, ya que una decisión que no puede ser explicada difícilmente puede ser justificada y aplicada. Esto implica que tanto los datos utilizados como los procesos que generan las decisiones deben ser trazables y estar documentados de manera clara.

La explicabilidad en IA significa que las decisiones que toma el sistema deben poder ser entendidas por aquellas personas que lo utilizan o se ven afectadas por sus decisiones. En el ámbito de la salud, esto es clave para que médicos y pacientes comprendan cómo y por qué se llegó a un diagnóstico o recomendación. Sin embargo, a veces hay que hacer equilibrios entre hacer el sistema más fácil de explicar, lo que puede reducir su precisión, o hacerlo más pre-

ciso pero más difícil de entender. Siempre que un sistema de IA impacte significativamente en la vida de las personas, debe ser posible exigir una explicación clara y accesible sobre cómo se tomaron esas decisiones, adaptada al nivel de conocimiento del usuario. Además, es crucial que pacientes y médicos sean informados cuando están interactuando con un sistema de IA para garantizar que no se confunda con un humano, lo que para ciertos grupos de población (personas mayores, poco diestras con las tecnologías, adolescentes, etc.) es crítico. Esto fomenta una interacción ética y clara entre los sistemas de IA y los usuarios, lo que aumenta la confianza en cuanto a su implementación y uso.

5. Diversidad, no discriminación y equidad

La no discriminación y equidad en los sistemas de IA es altamente relevante en el ámbito de la salud, donde es esencial que los sistemas sean inclusivos y accesibles para todos los pacientes, sin importar sus características individuales. Esto significa que se deben evitar los sesgos injustos que pueden surgir de datos históricos que perpetúan algún prejuicio específico, ya que esto podría llevar a discriminaciones en diagnósticos o tratamientos. Es vital eliminar estos sesgos desde la fase de recopilación de datos, para asegurar que las decisiones del sistema sean justas y transparentes.

Además, los sistemas de IA en salud deben diseñarse pensando en todos los pacientes, incluidos aquellos con discapacidades para asegurarse de que cualquier persona, sin que importe su edad, género o habilidades, pueda beneficiarse de estas tecnologías. También es fundamental involucrar a los diferentes actores del sistema de salud, incluidos pacientes, médicos y otros profesionales para garantizar que el desarrollo y la implementación de la IA sean equitativos y estén centrados en el bienestar de todos.

6. Bienestar social y ambiental

Este punto destaca la importancia de que los sistemas de IA sean sostenibles y respetuosos con el medio ambiente para

minimizar su huella energética. Además, la IA en salud debe promover la inclusión y fortalecer la cohesión social, garantizando que todos los grupos de la población puedan beneficiarse de los avances médicos. Asimismo, debe operar en consonancia con los valores democráticos, respetando los derechos de los pacientes y fomentando la confianza de la sociedad en su uso.

7. Responsabilidad

Los resultados generados por los sistemas de IA pueden afectar directamente al diagnóstico, tratamiento y bienestar de los pacientes. Para garantizar una clara rendición de cuentas, es fundamental que se definan líneas de responsabilidad a lo largo de todo el ciclo de vida del sistema, desde su diseño hasta su implementación y uso clínico. La auditabilidad desempeña un papel clave en este proceso, pues permite evaluar los algoritmos, los datos clínicos utilizados y los procesos de diseño y asegura que todo sea transparente y trazable, incluso cuando haya cuestiones de propiedad intelectual. También es importante identificar, documentar y minimizar los posibles impactos negativos que la IA pueda tener en los pacientes, ya que cualquier fallo podría tener consecuencias graves. Por último, disponer de mecanismos de reparación cuando los sistemas no funcionen como se espera es crucial para mantener la confianza de pacientes y profesionales, asegurando que existan soluciones cuando surjan problemas.

Retos y oportunidades

Los retos de la IA confiable en salud abarcan varias áreas clave. Un desafío importante es lograr la explicabilidad de sistemas complejos, como los basados en aprendizaje profundo caracterizados por excelentes y precisos resultados, que son difíciles de entender por los profesionales de la salud. La falta de claridad y las limitaciones en la validación clínica de las explicaciones acerca de



“La supervisión humana es indispensable para interpretar correctamente las recomendaciones de la IA, asegurando que el juicio clínico siga siendo el pilar fundamental en el cuidado del paciente.”

las decisiones de la IA puede reducir la confianza de médicos y pacientes. Además, garantizar la seguridad de los datos y la privacidad de los pacientes es fundamental, evitando filtraciones y un mal uso de información médica sensible. Otro reto es asegurar que todos los actores involucrados, desde desarrolladores hasta usuarios finales de la IA (pacientes y médicos), aboguen por el cumplimiento de estos requisitos de IA confiable, lo cual requiere un esfuerzo conjunto.

Sin embargo, las tensiones que a veces aparecen entre estos requisitos, por ejemplo, entre la precisión y la transparencia, también son un desafío, ya que

no existe una solución única para todos los casos. Esto exige reflexiones cuidadosas para abordar los dilemas y garantizar que las tecnologías de IA en salud se utilicen de manera segura, equitativa y efectiva, maximizando sus beneficios y minimizando riesgos. Por último, aunque la IA puede ser una valiosa asistencia en la toma de decisiones clínicas, nunca debe verse como un sustituto del profesional médico. La supervisión humana es indispensable para interpretar correctamente las recomendaciones de la IA, asegurando que el juicio clínico siga siendo el pilar fundamental en el cuidado del paciente.

En conclusión, la IA confiable puede promover la equidad en la atención médica, reduciendo sesgos y garantizando que todos los pacientes reciban un tratamiento justo. Además, su transparencia y explicabilidad generan confianza entre médicos y pacientes, lo que facilita su adopción en la práctica clínica. Esto revertirá en múltiples beneficios donde la IA permitirá la personalización del tratamiento, ajustando las terapias a las necesidades individuales de cada paciente, optimizando los recursos sanitarios y acelerando la investigación médica, mejorando la gestión hospitalaria y facilitando el descubrimiento de nuevos tratamientos y medicamentos mediante el análisis de grandes volúmenes de datos.