Aspectos éticos y legales de la inteligencia artificial
Ethical and legal aspects of artificial intelligence

M. Baget Bernaldiz, J. Téllez Vázquez, P. Romero Aroca

Resumen

Los médicos y los pacientes aceptaremos la implementación de los algoritmos si, además de ser eficaces, son transparentes, protegen la privacidad de los datos de los pacientes, entendemos cómo generan sus resultados y queda establecida de quién es la responsabilidad de su funcionamiento. Si participamos como oftalmólogos en el proceso de creación y monitorización de los algoritmos que afecten a la salud ocular, facilitaremos que nuestros pacientes confíen en ellos.

Palabras clave: Algoritmo. Sesgos. Privacidad de los datos. Responsabilidad.

Resum

Tant metges com pacients, anirem acceptant la implementació dels algorismes si a més a més de ser eficaços, son transparents, protegeixen la privacitat de les dades dels pacients, entenem com generen els resultats i queda ben establerta de qui és la responsabilitat del seu funcionament. Si participem com a oftalmòlegs en el procés de creació i monitorització dels algorismes que afecten la salut ocular, facilitarem la confiança dels nostres pacients en ells.

Paraules clau: Algorisme. Biaixos. Privacitat de les dades. Responsabilitat.

Abstract

Both patients and doctors will accept the presence of algorithms in our daily practice if, in addition to being effective, they are transparent, patient privacy is guaranteed, we understand how they generate the results and overall, it is well established who is responsible for their performance. If we, as ophthalmologists, participate in the process of creating and monitoring them, it will be easier for our patients to trust them.

Key words: Algorithm. Biases. Data privacy. Liability.

6.2. Aspectos éticos y legales de la inteligencia artificial

Ethical and legal aspects of artificial intelligence

M. Baget Bernaldiz[1,2,3], J. Téllez Vázquez[5,6], P. Romero Aroca[2,3,4]

[1]Médico adjunto del Servicio de Oftalmología. Hospital Universitari Sant Joan. Reus. Tarragona. [2]Universitat Rovira i Virgili. Reus. Tarragona. [3]Institut d'Investigació Sanitària Pere Virgili (IISPV). Reus. Tarragona. [4]Jefe del Servicio de Oftalmología. Hospital Universitari Sant Joan. Reus. Tarragona. [5]Médico adjunto del Servicio de Oftalmología. Hospital de la Santa Creu i Sant Pau. Barcelona. [6]Centro de Oftalmología Barraquer. Barcelona.

Correspondencia:
Marc Baget Bernaldiz
E-mail: mbaget@gmail.com

Introducción

La implementación de la inteligencia artificial (IA) va a generar un cambio de paradigma en la mayoría de los aspectos de nuestras vidas. Como oftalmólogos, tomaremos decisiones basadas, al menos en parte, en algoritmos que nos asistirán en el cribado, el diagnóstico y el tratamiento de las enfermedades oculares más prevalentes. A su vez, los pacientes deberán cuestionarse de qué manera aceptarán que este nuevo actor condicione su salud ocular.

Como ciudadanos, deberemos combatir la desinformación con un mayor sentido crítico. Las *fake news* (noticias falsas) son cada vez más elaboradas y difíciles de detectar, debido al uso de las redes neuronales generativas adversarias que generan vídeos falsos indistinguibles de los verdaderos.

Las empresas tecnológicas se han lanzado a la carrera para ver cuál de ellas saca al mercado el mejor algoritmo en su disciplina, ya sea un *chatbot* o un algoritmo médico. Esto puede ocasionar problemas de privacidad en los datos de los pacientes, así como la creación de algoritmos que no sean suficientemente precisos una vez aplicados en el mundo real.

Quizás ha llegado el momento de parar y reflexionar sobre hasta qué punto queremos que nos condicionen los algoritmos, de lo contrario, podría darse la situación de que los perjuicios pesen más que los beneficios. El documento recientemente firmado por un grupo relevante de profesores de universidad, investigadores y personas relevantes dentro de las empresas tecnológicas, entre ellas el propio Elon Musk, que pide parar el desarrollo de aquellos modelos de IA más potentes que el Chat GPT-4, es una muestra de ello. Aducen que se está produciendo una competición desmesurada entre las diferentes compañías de IA con la generación de algoritmos muy potentes, pero sin acabar de entender cuál es su funcionamiento y, por lo tanto, impredecibles y fuera de control.

Si pretendemos disponer de algoritmos que aporten un beneficio en la salud ocular a escala global, debemos cuestionarlos, para luego decidir y legislar sobre: la propiedad y privacidad de

los datos de los pacientes, evitar la introducción de sesgos en los algoritmos[1], la ciberseguridad y, sobre todo, quién va a ser el responsable del funcionamiento de los algoritmos[2].

Aspectos éticos

Los profesionales sanitarios y los pacientes, junto con los responsables de la gestión sanitaria, debemos reflexionar sobre la manera en la que estamos dispuestos a implementar la IA en nuestro trabajo, con el objeto de convertirla en una herramienta útil, y así beneficiar a la salud de nuestros pacientes. Los aspectos éticos principales a tener en cuenta son: la seguridad y eficacia, la transparencia, disponer de algoritmos imparciales y libres de sesgos, la privacidad de los datos y el consentimiento informado (Tabla 1).

Seguridad y eficacia de los algoritmos

La compañía IBM creó la herramienta de IA *Watson for Oncology* con el objeto de recomendar el tratamiento más adecuado para los pacientes oncológicos. En el año 2018, se descubrió que algunas de las recomendaciones que había dado eran incorrectas. La causa que se atribuyó a dicho error fue haber utilizado datos imprecisos en el entrenamiento del algoritmo. IBM ocultó el error por más de un año[3].

De lo anterior se deduce que la predictibilidad de un algoritmo dependerá, en primer lugar, de la calidad de los datos que se han utilizado para entrenarlo y de la población donde se aplique. Si se pretendiera realizar un cribado de cáncer de piel en la población asiática mediante un algoritmo que ha sido previamente entrenado utilizando datos de población caucásica, este vería disminuida su precisión.

Ya existen en el mercado algoritmos validados para la realización del cribado automático de la retinopatía diabética (RD). Todos ellos declaran una precisión muy alta en la detección de la RD referible (RD moderada o superior). No obstante, existen publicaciones independientes donde se observa una disminución significativa de su eficacia cuando se les somete al escrutinio de otras poblaciones diabéticas que no se han utilizado para su entrenamiento[4].

La eficacia y seguridad de un modelo algorítmico es el primer requisito que se le exige para ser incorporado dentro de un sistema de salud. Una manera de aumentar la confianza en los algoritmos es poder participar activamente en su creación y monitorización, tal y como recomienda la Comisión Europea (CE)[5].

Transparencia

Lo ideal sería que los responsables en desarrollar algoritmos proporcionaran información respecto a los datos poblacionales que han utilizado para su creación, así como la tecnología empleada de manera abierta. No obstante, a menudo, las compañías tecnológicas se oponen a causa de los derechos de la propiedad intelectual. Una posible solución sería la auditoría de estas empresas tecnológicas por parte de un organismo gubernamental.

Un aspecto que preocupa a ingenieros, médicos y usuarios, son los algoritmos *black box* (caja negra). Son aquellos algoritmos que nadie sabe muy bien cómo toman internamente sus decisiones. Este fenómeno es característico de las redes neuronales y determina que no podamos entender ni, por lo tanto, explicar a cada uno de los pacientes la razón subyacente por la que el algoritmo les hace determinada recomendación. Mientras se investiga este

	Problema	Solución
Eficacia	Demostrada en poblaciones restringidas	Realización de estudios randomizados
Seguridad	No hay criterios de comparación entre algoritmos	Crear *datasets* públicos para comparar algoritmos
Sesgos	Ampliamente arraigados en la sociedad	Algoritmos compensados por raza, sexo…
Transparencia	No hay declaración de procedencia de la fuente datos	Creación de organismos de control públicos
Privacidad	Falta de privacidad de datos de los pacientes	Aplicar la legislación de manera restrictiva
Responsabilidad	No hay legislación clara específica para inteligencia artificial	Establecer un marco jurídico claro

Tabla 1. Aspectos éticos y legales sobre el desarrollo de algoritmos basados en la inteligencia artificial en salud.

punto, deberemos exigir la existencia de estudios randomizados que demuestren la efectividad y seguridad de los algoritmos respecto a los métodos *gold standard* (estándar de oro)[6].

Algoritmos imparciales y libres de sesgos

En el año 2014, ingenieros de Amazon desarrollaron un modelo de IA para que "surfeara" la red en busca de nuevos talentos para la empresa. En el año 2017, Amazon clausura el algoritmo por haber demostrado ser discriminatorio contra las mujeres. Para su creación, se utilizaron datos de años anteriores, donde todavía había un predominio de hombres entre los ingenieros, al igual que los que desarrollaron el algoritmo. El resultado fue una aplicación donde la contratación de mujeres estuvo injustamente discriminada[7].

En Florida (Estados Unidos), el algoritmo COMPAS se construyó para predecir el riesgo de los delincuentes a reincidir[8]. Mostró una clara discriminación hacia los delincuentes de raza negra, a pesar de que, en muchos casos, el historial delictivo de los individuos blancos era peor. El motivo volvió a ser que los datos utilizados para el entrenamiento del algoritmo estaban fuertemente sesgados.

De lo anterior se deduce que, si quisiéramos construir un algoritmo que leyera y clasificara las retinografías de los pacientes diabéticos con precisión, deberíamos entrenarlo con un número parecido de retinografías de cada tipo de RD y tendría que estar equilibrado en cuanto a edad, raza y sexo. Una vez entrenado y testado, habría que ser prudentes y aplicarlo en poblaciones de pacientes diabéticos parecidas a la muestra de entrenamiento[9].

Debido a que las compañías tecnológicas no ofrecen información detallada sobre el desarrollo de sus algoritmos, deberían crearse organismos públicos que velen por la ausencia de sesgos en los modelos de IA antes de poder aplicarse en el mundo real[10].

Privacidad de los datos de los pacientes

En el año 2017, el Royal Free NHS Foundation Trust ofreció datos de más de 1,5 millones de pacientes a la empresa Google DeepMind, para que testara una aplicación denominada Streams para la detección del fallo renal. Los pacientes no fueron informados de que sus datos serían utilizados para dicho uso[11]. La preguntas que se plantean son: ¿a quién pertenecen los datos? y ¿quién puede utilizarlos?

En términos monetarios, los datos valen mucho dinero. Las empresas tecnológicas han de entender que, sin los datos de los pacientes, no pueden generar ningún tipo de tecnología y, por lo tanto, deben corresponder a los pacientes de algún modo. Siguiendo con el ejemplo anterior, el Royal Free NHS Foundation Trust llegó a un acuerdo con GoogleMind mediante el cual, a cambio de poder disponer de los datos, los pacientes podían utilizar la aplicación durante cinco años de manera gratuita. Mediante este acuerdo, se añadió valor a los datos de los pacientes[12].

Los datos referentes a la salud de las personas son muy sensibles. Un uso malicioso de los mismos podría llegar a afectar sus oportunidades laborales, sus relaciones personales y la contratación de las aseguradoras[13]. Una solución sería aplicar a los datos clínicos la misma legislación altamente restrictiva que se aplica en los datos genéticos y biométricos[14].

Un aspecto relevante y no resuelto es si el paciente tiene derecho a eliminar su información de una base de datos cuando ya se ha utilizado para la construcción de un modelo de IA[13].

El consentimiento informado en la era de la inteligencia artificial

La irrupción de la IA afectará a nuestros protocolos de cribado, de diagnóstico y de tratamiento actuales y, por lo tanto, el tipo de relación bilateral y sin intermediarios que hemos mantenido hasta ahora con el paciente. Ello condicionará cambios significativos en el consentimiento informado.

Se plantea la cuestión de hasta qué punto los médicos tendremos la responsabilidad de educar a los pacientes respecto a la complejidad de la IA en general y del algoritmo que les afecte en particular, así como de la posibilidad de explicar posibles sesgos y fallos del sistema. No obstante, y siendo honestos, los médicos debemos asumir que, hoy por hoy, no podemos justificar las recomendaciones del algoritmo de manera precisa. Esto no quiere decir que no los vayamos a utilizar.

Al igual que existen medicamentos de los que desconocemos los mecanismos exactos de actuación, pero que los utilizamos porque previamente han demostrado su eficacia y seguridad en estudios prospectivos randomizados, así deberemos proceder con

los algoritmos. Sin la demostración previa de su eficacia y seguridad, ni los médicos ni los pacientes estaremos en disposición de incorporarlos en nuestro trabajo. Y el consentimiento informado que acabe aplicándose en nuestro país o en el marco de la Unión Europea (UE) deberá incluir este aspecto.

Aspectos legales

Dos aspectos cruciales en la implementación de la IA son definir quién es el responsable del funcionamiento de los algoritmos y cómo se debe proceder para velar por la privacidad y seguridad de los pacientes.

Responsabilidad del funcionamiento de los algoritmos

Un algoritmo que intervenga en cualquiera de las fases de cribado, evaluación, diagnóstico o tratamiento del paciente, se equipara desde un punto de vista legal a un aparato médico, y se denomina *software* de dispositivos médicos (MDSW, *medical device software*) dentro del marco legislativo de la UE.

Los MDSW se clasifican en función de la repercusión en la salud del paciente que supondría un fallo del sistema. Los de clase I no supondrían perjuicios en términos de salud, y los de clase III podrían poner en riesgo la vida del paciente. Los sistemas de cribado ocular para la RD son de clase IIa, porque si bien su fallo puede empeorar la visión, no suponen un riesgo a nivel de la salud general de los pacientes.

La CE no ha definido un marco de responsabilidad para la IA y la robótica. Declara que la legislación actual es válida para la regulación de las nuevas tecnologías. No obstante, en noviembre del 2019, se constituye el Grupo Experto Independiente en Responsabilidad y Nuevas Tecnologías, que emite un documento sobre IA y el internet de las cosas (IoT, *internet of things*), responsabilizando a cada país de la UE el establecimiento de dicho marco legal, excepto para algunas cuestiones estrictas que serán competencia de la UE. Por lo tanto, vemos cómo hoy en día no está del todo resuelta la cuestión sobre quién tiene la responsabilidad del funcionamiento de los algoritmos.

Privacidad de los datos de los pacientes

La CE exige, desde el año 2018, la protección de los datos de los ciudadanos en todos los estados miembros de la UE (*The General Data Protection Regulation*, artículo 99(2) del Reglamento General de Protección de Datos [RGPD]). También protege a los ciudadanos en el caso de que sus datos fueran manejados por empresas localizadas fuera de la UE.

Con respecto a los datos de los ciudadanos que hacen referencia a la salud, datos biométricos o genéticos, la legislación es más exigente. Se permite su uso solo cuando el ciudadano ha dado su consentimiento explícito o bien lo exigen las cuestiones de salud comunitaria.

Ciberseguridad

Los *hackers* nos demuestran a diario su gran habilidad para sustraer información sensible de hospitales a través de la introducción de virus maliciosos en los servidores. Tenemos ejemplos fuera y dentro de casa. En Inglaterra, el *National Health Service* (NHS) fue paralizado en el año 2018 por un *ransomware* (secuestro de datos), que acabó justificando una nueva ley de ciberseguridad por parte de la UE en el año 2019. Recientemente, el Hospital Clínic de Barcelona ha sufrido un ataque similar, en el que los datos de algunos pacientes han sido expuestos públicamente.

En un futuro próximo, el IoT, los algoritmos y todos los datos circulando por la nube supondrán un reto para la ciberseguridad de cada centro sanitario. Hoy en día, los modelos de IA son especialmente vulnerables[15]. Un ataque malicioso podría causar cambios en los datos de la salud de los pacientes o alterar la precisión de un algoritmo. Cualquiera de estas situaciones conllevaría un riesgo para la salud de los pacientes.

La ciberdelincuencia no conoce fronteras, por lo que los expertos en ciberseguridad piden establecer medidas a escala global para hacer frente a este reto.

Conclusión

La implementación de la IA en nuestro trabajo como oftalmólogos va a ser una realidad en un breve periodo de tiempo. Debemos exigir a las empresas tecnológicas el desarrollo de algoritmos eficaces y seguros. Nuestra implicación en los procesos de desarrollo y monitorización de los algoritmos aportará valor añadido a nuestra profesión y mayor confianza en los pacientes.

Bibliografía

1. Gerke S, Minssen T, Cohen G. Ethical and legal challenges of artificial intelligence-driven healthcare. En: *Artificial Intelligence in Healthcare*. Elsevier; 2020. p. 295-336.
2. Naik N, Hameed BMZ, Shetty DK, Swain D, Shah M, Paul R, *et al*. Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility? *Front Surg*. 2022;9:862322.
3. Ross C, Swetlitz I. *IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments, internal documents show*. [Internet]. En: Statnews.com. STAT+. 25 Jul 2018. Disponible en: https://www.statnews.com/wp-content/uploads/2018/09/IBMs-Watson-recommended-unsafe-and-incorrect-cancer-treatments-STAT.pdf
4. Lee AY, Yanagihara RT, Lee CS, Blazes M, Jung HC, Chee YE, *et al*. Multicenter, head-to-head, realworld validation study of seven automated artificial intelligence diabetic retinopathy screening systems. *Diabetes Care*. 2021;44(5):1168-75.
5. Parlamento Europeo. Council Directive 2013/59/Euratom of 5 December 2013. *Off J Eur Commun L13*. 2014.
6. London AJ. Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability. *Hastings Cent Rept*. 2019;49(1):15-21.
7. Weismann J. Amazon Created a Hiring Tool Using A.I. It Immediately Started Discriminating Against Women. En: Moneybox. 10 Oct 2018.
8. Fefegha A. Racial Bias and Gender Bias in AI systems. En: Medium.com. 2 Sep 2018.
9. Price II WN. Medical Ai and Contextual Bias. *Harv J Law Technol*. 2019;33(1):65-116.
10. Sharkey N. The impact of gender and race bias in AI. En: Humanitarian Law & Policy. [Blog]. International Committee of the Red Cross (ICRC). 28 Ago 2018.
11. Information Commissioner's Office. Royal Free - Google DeepMind trial failed to comply with data protection law. ICO. 2017.
12. Cohen IG, Lynch HF, Vayena E, Gasser U. Big Data, Health Law, and Bioethics: Introduction (Cambridge University Press, 2018). *SSRN Electronic Journal*. 2018.
13. Gerke S, Minssen T, Yu H, Cohen IG. Ethical and legal issues of ingestible electronic sensors. *Nature Electronics*. 2019;2:329-34.
14. Roberts JL, Cohen IG, Deubert CR, Lynch HF. Evaluating NFL player health and performance: Legal and ethical issues. *University of Pennsylvania Law Review*. 2017;165:227-314.
15. Finlayson SG, Bowers JD, Ito J, Zittrain JL, Beam AL, Kohane IS. Adversarial attacks on medical machine learning. *Science*. 2019;363(6433):1287-9.