ISSN 1669-9106 MEDICINA (Buenos Aires) 2026: 86: 859-860

LOS EQUÍVOCOS AL INTERPRETAR LA "SUPERIORIDAD DIAGNÓSTICA DE LA INTELIGENCIA ARTIFICIAL" Y LA INTELIGENCIA ARTIFICIAL EXPLICABLE

CARLOS J. REGAZZONI

Comité de Salud Global y Seguridad Humana, Consejo Argentino para las Relaciones Internacionales (CARI), Instituto de Salud Global, Universidad J. F. Kennedy, Buenos Aires, Argentina

E-mail: cregazzoni@gmail.com

Agradezco los comentarios del Dr. Laudanno[1], que abordan un debate central para la medicina, como es la relación entre poder predictivo, opacidad algorítmica y responsabilidad clínica, en referencia a herramientas de inteligencia artificial (IA). Coincido en que el problema no debe reducirse a una oposición simplista entre causalidad y predicción. Sin embargo, considero necesario precisar tres puntos en los que disiento con la carta del Dr. Laudanno: 1) contraponer el *primum non nocere* a la interpretación causal, 2) sobreestimar el alcance real de la inteligencia artificial explicable (XAI), y 3) su interpretación respecto de la supuesta "superioridad" diagnóstica de los modelos de IA.

El *primum non nocere* constituye, sin lugar a dudas, un principio ético fundamental de la práctica médica, pero no puede opacar la necesidad de comprensión causal. Para la ciencia médica, tanto beneficencia como no-maleficencia se fundamentan en el conocimiento causal para explicar mecanismos, y para identificar condiciones de validez y generabilidad. En términos epistemológicos, el éxito predictivo que pueda tener un sistema de IA pertenece a un plano distinto del de la justificación y la inteligibilidad de nuestras inferencias, necesariamente basadas en la causalidad e indispensables para actuar según nuestro criterio[2]. Subordinar la interpretación causal a la mejora de métricas de desempeño predictivo confunde el plano ético con el epistemológico[3]. Despejar la necesidad de causalidad de la práctica médica redundaría en la imposibilidad de generalizar incluso aquellos hallazgos predichos por los modelos más robustos, ya que la generalización finalmente descansa en la causalidad, por el problema de transportabilidad de los hallazgos de un experimento causal[4].

En segundo lugar, punto central, cuando se afirma que *"si una herramienta incrementa la precisión diagnóstica del 90% al 98%, el deber ético se inclina hacia la opción más eficaz"*, como lo hace el Dr. Laudanno en su carta, se presta a cierta confusión conceptual desde la teoría bayesiana del diagnóstico. Dichas formulaciones sugieren implícitamente la existencia de una precisión diagnóstica global, independiente del contexto clínico. Sin embargo, desde la perspectiva de Bayes[5], toda probabilidad diagnóstica es condicional; el valor informativo de una prueba depende necesariamente de la probabilidad pretest del evento, así como de su sensibilidad y especificidad. Y todo resultado provisto por un algoritmo de IA[6], incluso todo resultado provisto por un gran modelo predictivo, finalmente constituye una probabilidad diagnóstica, y es condicional. Con demasiada frecuencia se habla de la superioridad diagnóstica de la IA sin hacer referencia a este aspecto decisivo[7]. Para no mencionar la cuestión de la calibración de los grandes modelos de lenguaje de acuerdo a contextos y uso, una cuestión de la máxima relevancia, aunque ignorada con frecuencia[8]. Las limitaciones mencionadas aplican a los grandes modelos de lenguaje, y deberían prevenirnos al momento de evaluar los trabajos que sostienen superioridad diagnóstica[9].

La analogía con instrumentos diagnósticos clásicos resulta ilustrativa. El microscopio posee una capacidad resolutiva muy superior al ojo humano para detectar malignidad en una biopsia; sin embargo, ello no implica que el microscopio sea "superior" al médico. Lo correcto es afirmar que el médico asistido por una herramienta diagnóstica adecuada, integrada en un razonamiento bayesiano y contextual, es superior a un médico privado de ella. Del mismo modo, los sistemas de IA no reemplazan al juicio médico, sino que modifican las condiciones bajo las cuales dicho juicio actualiza probabilidades diagnósticas, como elaboré extensamente en mi artículo[3].

En tercer lugar, si bien comparto el interés respecto de la XAI, discrepo con la idea de que la opacidad algorítmica sea esencialmente un problema de diseño, finalmente solucionable. Muchas técnicas de XAI producen explicaciones *post hoc* que correlacionan entradas y salidas, pero no explican el proceso inferencial subyacente en términos causales. Como ha señalado Rudin, estas explicaciones pueden ser plausibles sin ser verdaderas, generando incluso una ilusión de comprensión[10]. Este límite se vuelve aún más relevante en los grandes modelos de lenguaje, donde la salida corresponde al valor esperado de una distribución probabilística sobre secuencias lingüísticas, sin semántica causal interna susceptible de ser explicada.

Por ello, sostengo que la opacidad algorítmica no es un defecto accidental, sino una consecuencia estructural de modelos altamente parametrizados entrenados para maximizar desempeño predictivo[11]. La integración responsable de la IA en la práctica médica no puede lograrse al precio de una renuncia epistemológica, sino partiendo de un conocimiento profundo de la misma. El desafío no es solo adoptar herramientas más precisas, sino comprender qué tipo de inferencia realizan, cómo actualizan probabilidades y cuáles son los límites bayesianos de su aplicabilidad clínica.

En este sentido, el problema no es causalidad *versus* predicción, sino qué tipo de medicina estamos dispuestos a practicar en la era de la automatización de la inferencia.

Bibliografía

1. Laudanno O. Inteligencia artificial en medicina: entre la opacidad algorítmica y la responsabilidad clínica. *Medicina (B Aires)* 2025; 85:XX.
2. Otsuka J. Causal inference. En: Thinking About Statistics: The Philosophical Foundations of Statistical Reasoning. London, UK: Routledge, 2023, cap. 5, p144.
3. Regazzoni C. Inteligencia artificial y la era de las soluciones médicas inexplicables: navegando la opacidad algorítmica en la medicina actual. *Medicina (B Aires)* 2025; 85: 1076–92.
4. Pearl J, Bareinboim E. External validity: from do-calculus to transportability across populations. *Statist Sci* 2014; 2: 579-95.
5. Bours MJ. Bayes' rule in diagnosis. *J Clin Epidemiol* 2021; 131:158-60.
6. Hunter DJ, Holmes C. Where medical statistics meets artificial intelligence. *N Engl J Med* 2023; 389: 1211-9.
7. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019; 17: 195.
8. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Topic Group. 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: The Achilles heel of predictive analytics. *BMC Med* 2019; 17: 230.
9. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med* 2023; 29: 1930-40.
10. Rudin C. Stop explaining black box machine learning models for high stakes decisions. *Nat Mach Intell* 2019; 1: 206-15.
11. Lipton ZC. The mythos of model interpretability. *Queue* 2018; 16: 31–57.