IA y sesgos en salud

Nicolás Saá Muñoz – Académico del Instituto de Bioética de la Facultad de Medicina de la UCSC

¿Puede una inteligencia artificial profundizar las injusticias que pretende resolver? Los estudios en salud obligan a responder que sí. No porque la máquina "quiera" discriminar, sino porque aprende de sistemas que distribuyen de manera desigual la atención y los recursos.

En evaluaciones publicadas entre 2025 y 2026, distintos modelos de inteligencia artificial fueron enfrentados a casos clínicos idénticos, modificando solo variables como ingreso, vivienda, raza, edad u orientación sexual. Los resultados fueron inquietantes. Pacientes de mayores ingresos recibieron con frecuencia recomendaciones de tomografía o resonancia, mientras que aquellos descritos como pobres fueron derivados a estudios más básicos o a no realizar nuevos exámenes.

Algo semejante ocurrió en simulaciones de asignación de recursos escasos. La edad, los ingresos y el número de cargas familiares influyeron en la prioridad otorgada por la IA. La necesidad clínica dejó de ser el único criterio y comenzó a mezclarse con factores sociales, los cuales determinaban quién recibía una oportunidad y quién no.

Estos hallazgos actualizan la advertencia de Ziad Obermeyer. En su estudio, un algoritmo confundía gasto sanitario con necesidad de atención. Hoy, algunos sistemas parecen confundir pobreza con menor acceso a tecnología diagnóstica, vulnerabilidad con sospecha psiquiátrica y juventud con mayor prioridad en la asignación de recursos.

La dificultad no consiste solo en que la IA pueda equivocarse, sino en que sus errores no se distribuyen de forma equitativa. Un algoritmo puede ser preciso en promedio y, sin embargo, fallar frente a quienes ya enfrentan barreras históricas o han

sido postergados por la sociedad.

Desde una bioética personalista, esto resulta grave. La vulnerabilidad debería convocar mayor cuidado, no transformarse en una desventaja algorítmica. La persona no puede ser reducida a una correlación ni su dignidad depender de su edad, ingreso, orientación sexual, vivienda o pertenencia social.

El riesgo aumenta cuando estas herramientas se incorporan sin validación local, transparencia ni supervisión humana. Una decisión discutible puede adquirir apariencia de objetividad y convertir desigualdades previas en recomendaciones difíciles de cuestionar.

Antes de incorporar una IA a la Medicina como herramienta sin supervisión humana, no basta con preguntar cuánto acierta. Debemos preguntar para quién se equivoca, qué daño produce ese error y si la persona afectada puede ser escuchada y reparada dada la injusticia sanitaria-social cometida. Por lo que una tecnología que aprende de una sociedad injusta puede terminar haciendo más eficiente hasta la misma desigualdad que quiere erradicar.