Gac Sanit. 2026;40(S4):109

Gaceta Sanitaria

https://www.gacetasanitaria.org

455 - SESGOS COGNITIVOS EN LA INTELIGENCIA ARTIFICIAL PARA LA SALUD: UNA REVISIÓN DE ALCANCE SOBRE MARCOS DE GOBERNANZA Y ESTRATEGIAS DE MITIGACIÓN

S. Amaya-Santos, C. Vargas-Campos, C. Bermúdez-Tamayo

Universidad de Granada; Escuela Andaluza de Salud Pública; Ibs.Granada.

Resumen

Antecedentes/Objetivos: El despliegue de la Inteligencia Artificial (IA) en el sector salud ha priorizado históricamente la corrección de errores algorítmicos, dejando en un segundo plano los sesgos cognitivos. Estos errores sistemáticos en el pensamiento humano afectan la interpretación de las recomendaciones de la IA, pudiendo derivar en decisiones clínicas erróneas. Esta revisión analiza cómo los marcos de gobernanza actuales integran soluciones para gestionar la interacción cognitiva entre el profesional y la tecnología.

Métodos: Se llevó a cabo una revisión de alcance bajo el marco de Arksey y O'Malley y las guías PRISMA-ScR. La búsqueda abarcó seis bases de datos científicas (incluyendo PubMed, Embase e IEEE Xplore) analizando marcos de IA publicados entre enero de 2014 y marzo de 2025. El estudio evaluó cómo estos marcos abordan el sesgo a lo largo del ciclo de vida del sistema (diseño, validación, implementación y monitoreo).

Resultados: Se identificó que la gran mayoría de las directrices se concentran en la fase de desarrollo técnico (88,2%), con una atención notablemente menor en la implementación (58,8%) y el monitoreo posdespliegue (35.3%). Solo un 23.5% de los marcos analizados proponen estrategias específicas para el sesgo cognitivo. Entre las soluciones emergentes identificadas, destacan: 1) el diseño de interfaces de usuario explicables que eviten el exceso de confianza, 2) protocolos de "desbloqueo cognitivo" mediante alertas críticas, y 3) programas de capacitación en alfabetización algorítmica. No obstante, la operacionalización de estas estrategias sigue siendo teórica en la mayoría de los casos.

Conclusiones/Recomendaciones: Existe una desconexión crítica entre la teoría ética y la práctica clínica respecto al sesgo de automatización y otros errores cognitivos. Para garantizar una IA verdaderamente segura, los marcos de gobernanza deben evolucionar desde soluciones puramente técnicas hacia estrategias sociotécnicas. Es fundamental que el diseño de los sistemas de IA incorpore activamente mecanismos que desafíen el juicio del clínico en lugar de simplemente confirmarlo, asegurando así una toma de decisiones responsable y supervisada.