



## 6101-16. EVALUANDO LA CAPACIDAD DE LA INTELIGENCIA ARTIFICIAL EN LA ADECUACIÓN DE INFORMES CLÍNICOS PARA LA ALFABETIZACIÓN EN SALUD DEL PACIENTE CARDIOLÓGICO

Sarai Suárez Sampedro<sup>1</sup>, María Calvo Barceló<sup>2</sup> y Marco Tomasino<sup>2</sup>

<sup>1</sup>Hospital Universitari Vall d'Hebron, Barcelona, España y <sup>2</sup>Servicio de Cardiología. Hospital Universitari Vall d'Hebron, Barcelona, España.

### Resumen

**Introducción y objetivos:** La comunicación médico-paciente enfrenta el desafío de que los pacientes comprendan los informes médicos. En medicina, el principio de adecuación resalta la importancia de adaptar la información y la comunicación al nivel y contexto del paciente. La inteligencia artificial generativa (IAG) posee competencias lingüísticas avanzadas. El objetivo de este estudio es analizar y medir la capacidad de diversos modelos de lenguaje avanzado (LLM) de código abierto y propietario, para interpretar informes de alta médica y generar un reporte más accesible y comprensible para el paciente cardiológico.

**Métodos:** En el estudio se evaluaron 3 modelos de IA LLM (GPT-4; Claude y Llama2). Se generaron 6 informes IA a partir de 2 informes de alta médica del servicio de cardiología, anonimizados. Dos expertos en informes de alta del servicio realizaron la evaluación, que consta de un cuestionario con tres secciones: a) Instrumento Physician Documentation Quality Instrument (PDQI-9) con 9 ítems (actualización, precisión, minuciosidad, utilidad, organización, comprensibilidad, brevedad, concisión y consistencia del informe) y escala Likert 5. b) 10 Preguntas sobre concordancia, comprensibilidad, memorabilidad, empatía, alucinaciones, omisión de información, sesgo, riesgo de daño y legislación con escala Likert 5 y c) tres preguntas abiertas.

**Resultados:** Los informes generados por Claude y GPT-4 mostraron una concordancia clínica similar al original, pero con mayor claridad y comprensibilidad. Demostraron eficacia para simplificar textos y terminología, mejorando la memorabilidad al estructurar mejor la información de tratamiento en tablas (GPT-4). También recibieron puntuaciones altas en empatía. Se pudo comprobar que no incluyeron información falsa, aunque se identificaron omisiones importantes, como la historia cardiológica en los tres modelos. Solo el modelo Llama2 mostró riesgo de daño al omitir medicación. En cuanto a la calidad de documentación, PDQI-9, Claude y GPT-4 obtuvieron resultados similares al informe original. Las limitaciones del estudio incluyen un tamaño de muestra reducido y falta de evaluación con pacientes.

Análisis comparativo de herramientas de IA en la Interpretación de informes clínicos para pacientes

Aspecto evaluado	Claude	GPT-4	Llama2
------------------	--------	-------	--------

Instrumento PDQI-9 (máx. 45)	40	36,5	23,5
Concordancia (máx. 5)	4,5	4	2
Claridad (máx. 5)	4,5	4,25	3
Comprensibilidad (máx. 5)	4,5	4,25	2,5
Memorabilidad (máx. 5)	4,5	4,5	3
Empatía (máx. 5)	4,5	4	3
Alucinaciones (máx. 5)	1,25	1,5	3
Omisión de información (máx. 5)	3,5	2,5	5
Sesgo (máx. 5)	1	1	1
Información con riesgo de daño (máx. 5)	1	1,25	3,25
Cumplimiento de legislación (máx. 5)	3,5	3,25	2,25

Comparación de los 3 modelos de inteligencia artificial evaluados por Physician Documentation Quality Instrument (PDQI-9) y cuestionario *ad hoc*.

**Conclusiones:** Las IA Claude y GPT-4 simplifican informes médicos con detalle adecuado, mejorando la comprensión, memorabilidad y empatía en la comunicación con el paciente. Es imprescindible la supervisión profesional en aquella información con riesgo de daño y garantizar precisión y ética.