



Cirugía Española



www.elsevier.es/cirugia

O-148 - VALIDACIÓN DE LA HERRAMIENTA CHATGPT PARA LA EXTRACCIÓN DE DATOS EN HISTORIAS CLÍNICAS ELECTRÓNICAS DE CIRUGÍA BARIÁTRICA

Trinidad, Anna; Lobo, Nikolak; Paris, Marta; Bartra, Elia; Vives, Margarida; Raga, Esther; Sabench, Fatima; del Castillo Déjardin, Daniel

Hospital Universitari Sant Joan, Reus.

Resumen

Introducción: La extracción de datos de historias clínicas electrónicas (HCE) es un área en constante evolución que ha sido objeto de estudio en los últimos años. Los sistemas de procesamiento del lenguaje natural (PLN) han demostrado ser una herramienta útil para la extracción de información relevante de las HCE. Uno de estos sistemas es el modelo de lenguaje de gran escala (LLM) ChatGPT. El rendimiento de estos sistemas de realizar la extracción de datos de HCE de cirugía bariátrica aún no ha sido explorada.

Métodos: Se utilizó el modelo ChatGPT para la extracción de información de 10 HCE de cirugía bariátrica en castellano y catalán, incluyendo 9 variables: 2 continuas -índice de masa corporal y edad- y 7 categóricas-presencia o ausencia de hipertensión arterial, diabetes, dislipemia, antecedentes quirúrgicos, medicación, presencia de atracones, hábito tabáquico y hábito enólico-. La validación se ha llevado a cabo mediante la comparación de la extracción de datos realizada por el modelo con los datos extraídos manualmente por un especialista en cirugía bariátrica, que se utilizaron como *gold standard*. "Verdadero positivo" se definió como ambos resultados idénticos, "falso negativo" como la presencia del resultado de la variable en la revisión manual pero la ausencia de un puntaje extraído por la herramienta y "falso positivo" como cualquier resultado extraído por la herramienta pero sin resultado en la revisión manual, o el resultado que fue extraído por la herramienta pero no coincidió con el de la revisión manual. Se compararon los resultados obtenidos por el modelo con los datos de referencia mediante una matriz de confusión, para el cálculo de sensibilidad, el valor predictivo positivo (VPP) y F1 score.

Resultados: Los resultados obtenidos en la validación revelaron que hubo 77 datos considerados como verdaderos positivos, 23 falsos positivos y 1 solo caso de falso negativo. Estos datos dieron unas métricas la extracción de datos de HCE de cirugía bariátrica en castellano y catalán con una sensibilidad del 98%, un VPP del 77%, y un F1-score de 86%. Sin embargo, al equiparar la detección de casos ausentes y negaciones como ausencia de condición, se observó una mejora significativa en todas las métricas, con una sensibilidad del 99%, un VPP del 95% y un F1 score del 97%.

Conclusiones: El modelo ChatGPT es una herramienta útil para la extracción de información de HCE de cirugía bariátrica. Aunque se identificaron algunos desafíos, como una cifra algo elevada de FP, al equilibrar los casos de *missing* y negaciones, se observó una mejora significativa en todas las métricas. Estos resultados sugieren que el modelo ChatGPT tiene el potencial de mejorar la eficiencia y precisión en la extracción de información clínica de las HCEs en el futuro.