

## 70 - UTILIDAD DE LA I.A. EN LA DETECCIÓN DE ERRORES EN HISTORIAS CLÍNICAS

**Luis Jesús del Peral Rodríguez<sup>1</sup>**, Marta Torres Arrese<sup>2</sup>, Alma Real Martín<sup>2</sup>, Begoña Gonzalo Moreno<sup>2</sup>, Juan Gabriel Sánchez Cano<sup>2</sup>, Álvaro González-Blanch Rodríguez<sup>2</sup>, Ana Yi-Shi Wu Wang<sup>2</sup> y Davide Luordo Tedesco<sup>1</sup>

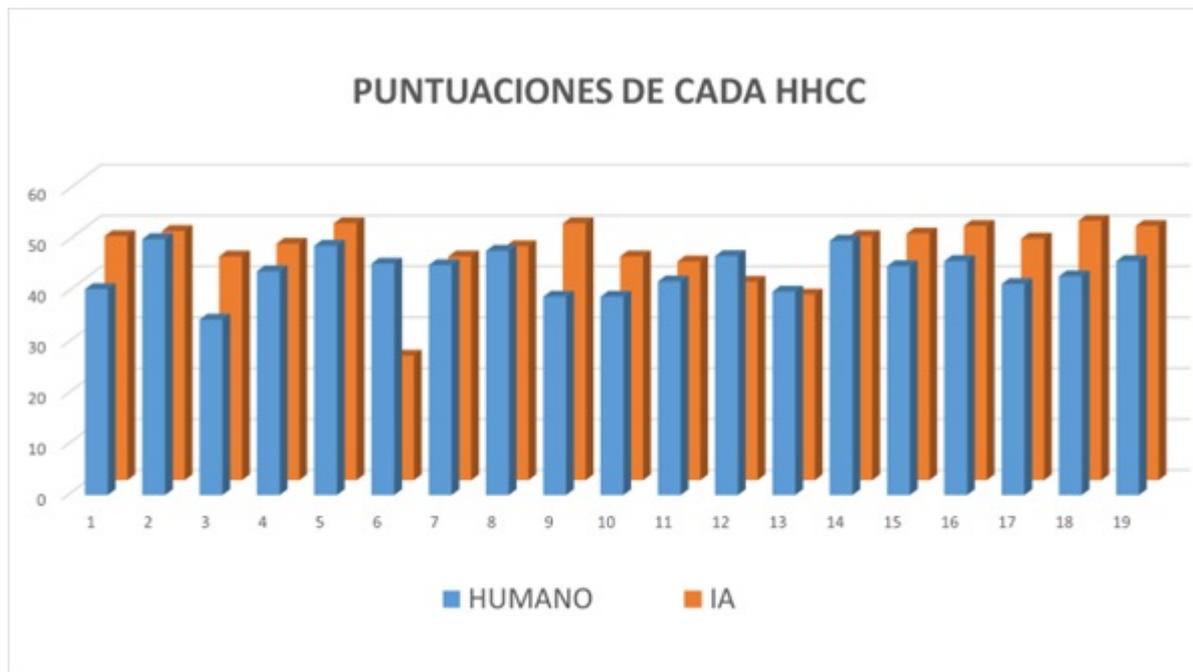
<sup>1</sup>Hospital Universitario Infanta Cristina, Parla, España. <sup>2</sup>Hospital Universitario Fundación Alcorcón, Alcorcón, España.

### Resumen

**Objetivos:** Los avances recientes en Inteligencia Artificial (IA), y la creación y apertura al público de diversas IA de tipo conversacional como ChatGPT o Google Bard especializadas en el procesamiento de lenguaje natural, han revolucionado otros ámbitos y profesiones. Nosotros proponemos que podrían ser útiles también para identificar incoherencias y errores en las historias clínicas (HHCC) de los pacientes.

**Métodos:** Se crearon 21 HHCC ficticias, entre las que se mezclaron errores de diagnóstico, de tratamiento y también incoherencias lingüísticas. Posteriormente fueron evaluadas según unos criterios determinados por ChatGPT 3.5 y por un médico independiente al proceso de creación de las HHCC. Cada historia podía recibir un máximo de 50 puntos. Dos historias tuvieron que ser desechadas por ser demasiado extensas para la IA. El análisis estadístico se realizó con SPSS 24.

**Resultados:** Las evaluaciones por el corrector humano y por IA para las respectivas 19 HHCC, pueden verse de forma gráfica en la imagen. La puntuación media de ambos grupos es similar (44 vs. 45) y el test de Wilcoxon no permite demostrar diferencias estadísticamente significativas entre ambos grupos. Sin embargo, la desviación estándar del corrector humano es bastante menor (4,2 vs. 6,4) y su distribución se adecuaba a la normal, mientras que la IA no. Además, el coeficiente de correlación intraclase es de 0,1.



*Discusión:* Lo primero que salta a la vista, es que la evaluación por IA de las HHCC, pese a otorgar una puntuación media similar, muestra mayor disparidad y muy baja concordancia con la corrección humana, como objetiva el análisis de la varianza y el coeficiente de correlación intraclase. Sin embargo y para sorpresa nuestra, la mayor utilidad de la IA resultó ser que proporcionaba justificación para sus evaluaciones, y en varios casos sugería una manera de mejorar la historia clínica, o ideas que podría haber incluido para una mejor puntuación, demasiado extensas para incluirlas en este trabajo. Esto nos parece de grandísima utilidad, ya que puede hacer ver a un corrector humano detalles en los que no podría haberse fijado. Se debe tener en cuenta, que los resultados del estudio están limitados por el número limitado de casos, y porque bien advierte ChatGPT en sus respuestas, no está específicamente entrenada para la detección de errores en HHCC ni en la elaboración de un diagnóstico o tratamiento médico.

*Conclusiones:* La IA puede llegar a ser una herramienta útil a la hora de corregir textos como pueden ser las historias clínicas. Es cierto que la tecnología de hoy está muy lejos de sustituir la mente humana ya que la IA proporciona resultados demasiado impredecibles y discordantes como para poder tomar por veraces los resultados sin la supervisión de un experto en la materia. Sin embargo, el *feedback* y las sugerencias aportadas por la IA, pueden resultar muy positivas en detectar errores, que posteriormente tengan que ser revisados por el experto. Es muy posible que el entrenamiento de una IA específicamente en este campo tenga gran utilidad para homogeneizar y garantizar la calidad de las historias clínicas en la práctica médica.