

Título del Trabajo:

Estudio Comparativo sobre la Calidad de la Información para Pacientes Provista por Agentes de inteligencia artificial (IA) Generativa sobre Sillas de Ruedas Adecuadas

Título del Trabajo en Inglés:

Comparative Study on the Quality of Patient Information Provided by Generative Artificial Intelligence (AI) Agents on Appropriate Wheelchairs

Nombre: ALBERTO ISAAC

Apellidos: PEREZ SANPABLO

ORCID: 0000-0003-0550-928X

País de Residencia: MEXICO

Área de Investigación: TECNOLÓGICA

Institución a la que Pertenece: INSTITUTO NACIONAL DE REHABILITACION GILLERMO IBARRA IBARRA

Área de Adscripción: Laboratorio de Análisis de Movimiento e Ingeniería de Rehabilitación, Colegio de Ingenieros Biomédicos de México

Correo Electrónico: albperezinr@gmail.com

Datos de los(as) coautores(as) del Trabajo

Alicia Meneses Peñaloza

Malformaciones Congénitas, INR, MEXICO, aliciamp1@hotmail.com, 0000-0003-4042-6244

Palabras en Español:

Inteligencia Artificial, Información para el paciente, Legibilidad

Palabras en Inglés:

Artificial Intelligence, Patient information, Readability

Título del Trabajo:

Estudio Comparativo sobre la Calidad de la Información para Pacientes Provista por Agentes de Inteligencia Artificial (IA) Generativa sobre Sillas de Ruedas Adecuadas

Título del Trabajo en Inglés:

Comparative Study on the Quality of Patient Information Provided by Generative Artificial Intelligence (AI) Agents on Appropriate Wheelchairs

Área de Investigación:

Laboratorio de Análisis de Movimiento e Ingeniería de Rehabilitación, Colegio de Ingenieros Biomédicos de México

Introducción:

Los agentes de IA generativa están ganando relevancia en salud, con potencial para ofrecer respuestas en lenguaje natural a pacientes y profesionales. Evaluar la calidad de esta información es crucial para su eficacia y seguridad. La información sobre servicios de sillas de ruedas es esencial para mejorar el cuidado de los usuarios. Los prompts, que median la interacción, y la legibilidad son críticos para asegurar la comprensión. Esta investigación analiza la precisión y claridad de esta información. Estudios preliminares indican una baja calidad en la información sobre sillas de ruedas. Sin embargo, no se ha abordado la legibilidad, alternativas de interacción ni los agentes actuales.

Objetivo:

Evaluar la calidad, legibilidad, fiabilidad y efecto de la interacción en la información proporcionada por modelos de agentes actuales tanto de acceso libre (ChatGPT4mini, Gemini y Copilot) como por suscripción (ChatGPT4o).

Metodología:

Se seleccionaron cuatro agentes (ChatGPT4mini, ChatGPT4o, Gemini y Copilot) debido a su accesibilidad y popularidad. Se utilizó la herramienta EQIP para evaluar la calidad de las respuestas. Dos expertos certificados en el tema revisaron de manera independiente la precisión de las respuestas utilizando una hoja de cálculo de Excel. Se utilizaron prompts estandarizados tipo “zero” y avanzados integrando 14 categorías básicas. Las consultas se realizaron dos veces en agosto 2024 utilizando nuevas ventanas para evitar sesgos. Esta muestra fue calculada para una fiabilidad mínima y esperada de 0.7 y 0.99 respectivamente. Se calculó la fiabilidad test-retest y entre evaluadores mediante el coeficiente de correlación intraclase (ICC). Además, se midió la legibilidad de la información utilizando 11 métricas con Grammarly® para una intención informada y formalidad neutral. Se empleó la prueba de Kruskal-Wallis para analizar el efecto de los prompts de interacción y diferencias entre agentes.

Resultados:

Los agentes conversacionales demostraron una fiabilidad test-retest muy alta ($ICC > 0.99 \pm 0.01$, $p < 0.01$). La fiabilidad entre evaluadores fue sobresaliente ($ICC = 0.96 \pm 0.03$, $p = 0.01$). Los agentes presentaron diferencias estadísticamente significativas en calidad (total e identificación) y legibilidad (número de palabras, tiempo de lectura y escritura, longitud de oraciones y uso de palabras únicas). El

agente por suscripción sobresalió en términos de calidad (EQIP=75%) y legibilidad (94%) de la información. Gemini presentó el rendimiento más bajo con una calidad y legibilidad entre 43% a 53% menor respecto al mejor agente. No se encontraron diferencias significativas entre los prompts avanzados y los básicos. Todo esto apoya la existencia de limitaciones inherentes en los modelos y en los datos de entrenamiento utilizados.

Conclusiones:

Los prompts avanzados no mejoraron significativamente la calidad y legibilidad de la información, pero la puntuación promedio de calidad mejoró respecto a estudios anteriores. El agente de suscripción sobresalió por su buen rendimiento. Se confirmó la fiabilidad, así como la utilidad de la metodología y la importancia de continuar la investigación.