

Evaluación cualitativa de modelos de inteligencia artificial generativa para resolución de preguntas clínicas de rehabilitación infantil

DANIELA GARCÍA PALOMER¹, MAURICIO ARRETX SPOERER¹

ABSTRACT

Qualitative Evaluation of Generative Artificial Intelligence Models for Solving Rehabilitation Clinical Questions

Introduction: The use of artificial intelligence in the healthcare field has shown multiple potentials. Among them, using generative artificial intelligence models, such as ChatGPT, to support the creation of clinical documents like reviews, clinical guidelines, or protocols. The objective of this study is to evaluate and compare the responses to clinical questions defined by experts on two rehabilitation topics, provided by the most widely used generative AI models on the market, in order to analyze their role in the creation of clinical documents, such as clinical guidelines or protocols. **Material or Patients and Methods:** Qualitative descriptive study. Through prompts of various types of clinical questions, an expert evaluates and rates, using a scale design for this study, the responses provided by 5 generative artificial intelligence models: ChatGPT, Gemini, Claude, Perplexity, and a customized GPT. **Results:** All generative artificial intelligence models are capable of delivering well-structured and coherent responses, but with some shortcomings in their technical content and updates according to evidence-based medicine. ChatGPT received the highest ratings, using the study rating scale. **Discussion:** Generative artificial intelligence models can play a role in the creation of clinical documents, such as reviews, guidelines, and protocols, providing a quick and effective tool. However, each step in the creation of these documents should be supervised by a human expert to identify possible errors, hallucinations and protection of ethic and security issues.

Keywords: Artificial intelligence, protocol, guideline.

RESUMEN

Introducción: El uso de inteligencia artificial en el área de la salud ha demostrado tener múltiples potenciales. Entre ellos, utilizar los modelos de

¹Equipo Innovación, Gerencia de Infraestructura y Tecnología, Teletón. Chile.

Recibido: 30-10-2024
Aceptado: 16-12-2024

Correspondencia:
Daniela García Palomer
dgarcia@teleton.cl

inteligencia artificial generativa, como ChatGPT, para apoyar en la confección de documentos clínicos como revisiones, guías clínicas o protocolos. El objetivo de este estudio es evaluar y comparar las respuestas a preguntas clínicas definidas por expertos en temas referentes a rehabilitación, otorgadas por los modelos de IA generativa más utilizados en el mercado, para así analizar su rol en la confección de documentos clínicos, como guías clínicas o protocolos.

Material o Pacientes y Métodos: Estudio descriptivo cualitativo. A través de *prompt* de preguntas clínicas de distinto tipo sobre dos temáticas de rehabilitación, un experto evalúa y califica con una pauta de apreciación diseñada para este estudio las respuestas entregadas por 5 modelos de inteligencia artificial generativa: ChatGPT, Gemini, Claude, Perplexity y GPT customizado. **Resultados:** Todos los modelos de inteligencia artificial generativa son capaces de entregar respuestas bien estructuradas y coherentes, pero con algunas falencias en sus contenidos técnicos y actualización acorde a medicina basada en la evidencia. ChatGPT presenta las mejores calificaciones acorde a la pauta utilizada. **Discusión:** Los modelos de inteligencia artificial generativa pueden jugar un rol en la creación de documentos clínicos, como revisiones, guías y protocolos, aportando una herramienta rápida y efectiva. Sin embargo, cada paso de la creación de estos documentos debiese ser supervisado por un experto humano, para pesquisar posibles errores, alucinaciones y resguardo de los aspectos éticos y de seguridad.

Palabras clave: Inteligencia artificial, protocolo, guía.

Introducción

La inteligencia artificial (IA), a grandes rasgos, es la capacidad que tienen las máquinas para realizar tareas que requieren de inteligencia humana, tales como reconocimiento de lenguaje, toma de decisiones acorde a algoritmos, y aprendizaje¹. Producto de estas capacidades, la IA ha revolucionado el mundo de la medicina a través de transformar la forma en que se realiza, desde la aplicación de procesos, hasta implicancias directas en la atención clínica.

La IA generativa es una clase de tecnología de *machine learning* que puede entrenarse y aprende a generar nueva data a través del entrenamiento. Uno de los modelos de generación de contenido más conocidos es ChatGPT-4 (*Generative Pretrained Transformer*), pero también existen otros en el mercado. Las herramientas y aplicaciones de IA generativa pueden automáticamente aprender patrones y estructuras a partir de texto, imágenes, soni-

dos, animaciones, modelos u otros medios, para generar data nueva con características similares². Se utilizan para hacer búsquedas, escribir, crear modelos, códigos o arte, sin la intervención de una persona. Incluso, se le puede solicitar a la IA generativa que asuma roles específicos, para mejorar la calidad de sus respuestas como, por ejemplo, el rol de un experto en una temática en particular.

En el ámbito de la medicina, son múltiples los posibles usos y aportes que se han descrito para la IA. Entre ellos, cabe mencionar la generación de data y aumentación de data, creación de fármacos, *triage* de pacientes, telemedicina, revisión e interpretación de exámenes, obtención de diagnósticos acuciosos, creación de documentos clínicos o resúmenes de ya existentes, administración eficiente de procesos, y personalización de la salud a las necesidades de cada paciente³⁻¹⁶. Esto último se ve en el análisis de data para poder proporcionar tratamientos e intervenciones personalizadas². Incluso, hay artículos que describen

sus aportes en la formación y educación de futuros profesionales de salud, de los propios pacientes para mejorar su adherencia y estimular la prevención^{2,17}. También hay artículos que abordan sus posibles usos dentro de la consulta clínica, como el apoyo con la toma de notas en ficha clínica, y en el razonamiento clínico, a través de generar diagnósticos diferenciales y apoyar en la interpretación de resultados de exámenes¹⁸.

Por otra parte, dado el rápido avance e incorporación de la IA a la medicina, existen preocupaciones con respecto al marco médico legal en que debe regularse, sobre cómo debe integrarse su uso de manera apropiada a los servicios de salud y cómo los equipos de salud deben utilizar este tipo de herramientas, o sobre cómo medir su real impacto en los resultados de los pacientes³. También aparecen aprensiones con respecto a posibles *bias* o sesgos en los algoritmos que puedan acarrear discriminaciones, inequidad en la distribución de recursos y, en consecuencia, acentuar las diferencias². Por otra parte, hay estudios que describen las falencias de la IA en la protección, seguridad y confidencialidad de la información sensible de los pacientes, y el impacto que esto podría acarrear en la confianza en las instituciones de salud¹⁹. Por esta razón, es fundamental que cualquier incorporación del uso de IA debe hacerse de manera responsable e informada, siempre considerando los aspectos éticos y el marco regulatorio, e incorporando monitoreos continuos, evaluaciones y la preparación y capacitaciones de los equipos.

Con respecto a la revisión y creación de artículos científicos, como revisiones sistemáticas y guías clínicas, también hay artículos publicados que intentan objetivar los aportes del uso de IA. La IA puede aportar desde la búsqueda de los artículos científicos que podrían aportar al documento, incluido la traducción, creación de resúmenes y/o análisis de estos²⁰. Puede ayudar a crear ensayos de tópicos específicos y entrenar su estilo para que asemeje el utilizado en la literatura científica, e incluso apoyar en el análisis estadístico²¹. Utilizando *prompts* adecuados, se podrían sintetizar revisiones con los detalles de los avances de varios campos o crear

guías clínicas de patologías específicas en segundos^{20,22}. Sumado a lo anterior, la actualización de estos documentos también podría realizarse de manera más rápida y eficiente utilizando IA²³. Es importante señalar que, en el proceso de escritura del documento científico, es crítica la guía, supervisión y juicio humano de expertos en el tema, para asegurar la acuciosidad, coherencia y credibilidad del contenido, antes de enviarlo a publicación¹⁹. Al respecto, aún está en discusión cuál es la mejor forma de citar el uso de herramientas de IA en la creación de publicaciones científicas²⁰, además de otras alertas que se han detectado tales como: muchas veces la información que entregan los modelos de IA generativa, a pesar de estar bien escritas, están desactualizadas, fabricadas y/o sin referencias²⁴. También se menciona la poca transparencia en la metodología de búsqueda, el uso de referencias falsas, la dificultad para restringir el uso de fuentes para asegurar calidad, la contextualización adecuada, tener que crear y recrear *prompts* complejos, y, por último, las denominadas “alucinaciones” de la IA²⁵⁻²⁷ que corresponden al fenómeno cuando la AI genera una respuesta bien escrita y convincente, pero completamente falsa²⁸. Esto incluye la entrega de referencias falsas o la confección de artículos científicos fraudulentos, pero de apariencia auténticos²⁹.

El objetivo general de este estudio es realizar una evaluación cualitativa y comparativa de las respuestas a preguntas clínicas otorgadas por los modelos de IA generativa más utilizados en el mercado, para así analizar el rol de estas herramientas en la confección de artículos científicos, como por ejemplo guías clínicas. Dado que los autores trabajan en un centro de rehabilitación infantil, se seleccionarán temas referentes a rehabilitación.

Material o Pacientes y Métodos

Modelos de inteligencia artificial generativa

Para este estudio se seleccionaron 5 modelos de IA generativa. Cuatro de ellos por ser los más utilizados y reconocidos actualmente

y el quinto (GPT Teletón) para evaluar las herramientas de customización de OpenAI. A continuación, se describen los 5 modelos:

1. ChatGPT 4o: Es un modelo de lenguaje desarrollado por OpenAI y pertenece a la familia GPT (Generative Pre-trained Transformer). ChatGPT 4o (o por “onmi”) es multimodal, lo que significa que puede procesar tanto texto como imágenes.
2. Gemini Advanced: Es la versión mejorada y de pago de Gemini, la herramienta de inteligencia artificial que Google desarrolló para reemplazar a Google Bard.
3. Claude-3.5 Sonnet: Claude es una familia de modelos de lenguaje desarrollados por Anthropic que incluyen en tareas que involucran lenguaje, razonamiento, análisis y codificación, y Claude-3.5 Sonnet es uno de sus modelos.
4. Perplexity: Fue creado por un equipo de expertos que trabajaban en Google y se basa en un modelo que combina las capacidades de GPT-3 y GPT-4.
5. Open AI customizado (GPT Teletón): OpenAI ofrece la posibilidad de crear y entrenar la herramienta de ChatGPT para customizarla a los usos y necesidades específicas de sus usuarios. GPT Teletón fue creada con un foco en rehabilitación integral, centrada en el paciente y su familia, y basado en altos estándares de calidad científica. Para esto, se entrenó el modelo entregándole detalles del rol que esperábamos, especificando su perfil de experto en rehabilitación, usuario de medicina basada en la evidencia, y con un foco en el paciente y familia. Como material de entrenamiento, se le entregaron los protocolos de atención específicos utilizados en nuestro centro de rehabilitación, además de los artículos científicos de referencia de cada pregunta señalados en los *prompt*.

Estructura de los *prompt* utilizados

Se definieron 2 preguntas clínicas de temas de rehabilitación para evaluar los distintos modelos de IA generativa. El primero fue referente al diagnóstico precoz de la parálisis cerebral y el segundo se centró en la prescripción protési-

ca en lactantes con deficiencias congénitas de extremidad superior. Se seleccionaron estos dos temas, pues cuentan con artículos científicos de referencia y con respuestas claras, acotadas y medibles por el evaluador. Para cada pregunta clínica, se realizarán 3 niveles de preguntas:

1. Pregunta clínica simple.
2. Pregunta clínica + datos de entrenamiento (artículos científicos específicos). Estos artículos se seleccionaron por su relevancia en la temática y por contener la información necesaria para poder responder la pregunta solicitada.
3. Solicitud de confección de protocolo que responda pregunta clínica y que incluya las siguientes secciones: i. Población Objetivo; ii. Marco Conceptual; iii. Objetivo General; iv. Objetivos específicos; v. Instrumentos de evaluación; vi. Desarrollo del Procedimiento; vii. Recomendaciones; viii. Bibliografía.

Es importante señalar que, para realizar una comparación objetiva de cada modelo de IA generativa, se utilizó el mismo *prompt*, sin pruebas de ajuste, para cada una de las herramientas y se registraron las respuestas textuales. En la Tabla 1, se resumen los *prompt* utilizados en el estudio.

Análisis de las respuestas

Para analizar las respuestas, se le solicitó a un experto clínico en rehabilitación infantil, integrante del equipo investigador y sin conflicto de interés, que revisara cada una de las respuestas entregadas por los distintos modelos de IA generativa y las calificara. Este experto fue un médico especialista en Medicina Física y Rehabilitación. Para la calificación, se construyó una pauta de apreciación especialmente para este estudio, pues no se encontraron referencias previas en la literatura (Tabla 2). La pauta fue confeccionada por los autores e incluía la calidad del contenido técnico, la claridad de la respuesta, nivel de actualización y su concordancia con la medicina basada en la evidencia. En base a los puntajes de esta pauta, se calificaron las respuestas en 3 categorías: “Buena”, “Regular” o “Mala”.

Tabla 1. Resumen de los *prompt* utilizados en el estudio

Tema	Pregunta clínica simple	Pregunta clínica + datos de entrenamiento	Confección de protocolo
Diagnóstico precoz de parálisis cerebral	¿Cómo se hace el diagnóstico de Parálisis Cerebral en una persona lactante menor de 5 meses de edad?	Cómo se hace el diagnóstico de Parálisis Cerebral en una persona lactante menor de 5 meses de edad Para ello considera el artículo Novak I, Morgan C, Adde L, Blackman J, Boyd RN, Brunstrom-Hernández J, Cioni G, Damiano D, Darrah J, Eliasson AC, De Vries LS. Early, accurate diagnosis and early intervention in cerebral palsy: advances in diagnosis and treatment. JAMA pediatrics. 2017 Sep 1;171(9):897-907	Por favor, redacta un protocolo para responder la pregunta clínica de cómo se hace el diagnóstico de Parálisis cerebral en una persona lactante menor de 5 meses de edad. El documento debe incluir las siguientes secciones: i. Población Objetivo ii. Marco Conceptual iii. Objetivo General iv. Objetivos específicos v. Instrumentos de evaluación vi. Desarrollo del Procedimiento vii. Recomendaciones viii. Bibliografía
Prescripción protésica en deficiencias congénitas de extremidad superior	¿Cuándo se indica la primera prótesis en un niño con deficiencia transversal de antebrazo?	Por favor, explícame cuándo correspondería indicar la primera prótesis en un niño con deficiencia transversal de antebrazo. Para responder, bázate principalmente en los siguientes 3 artículos para elaborar tu respuesta: Huizing, Karin, Heleen Reinders-Messelink, Carel Maathuis, Mijna Hadders-Algra, and Corry K. Van Der Sluis. "Age at first prosthetic fitting and later functional outcome in children and young adults with unilateral congenital below-elbow deficiency: a cross-sectional study." <i>Prosthetics and orthotics international</i> 34, no. 2 (2010): 166-174 Le, Joan T., and Phoebe R. Scott-Wyard. "Pediatric limb differences and amputations." <i>Physical medicine and rehabilitation clinics of North America</i> 26, no. 1 (2015): 95-108. Krebs, David E., Joan E. Edelman, and Maureen A. Thornby. "Prosthetic management of children with limb deficiencies." <i>Physical therapy</i> 71, no. 12 (1991): 920-934	Por favor, redacta un protocolo para responder la pregunta clínica de cuándo se indica la primera prótesis en un niño con deficiencia transversal de antebrazo. El documento debe incluir las siguientes secciones: i. Población Objetivo ii. Marco Conceptual iii. Objetivo General iv. Objetivos específicos v. Instrumentos de evaluación vi. Desarrollo del Procedimiento vii. Recomendaciones viii. Bibliografía

Tabla 2. Pauta de apreciación utilizada para calificar respuestas

Categoría	Bueno*	Regular**	Malo***
Calidad del contenido técnico de la respuesta entregada, con respecto a los estándares actuales	2 pt	1 pt	0 pt
Claridad, coherencia y comprensión de la respuesta entregada	2 pt	1 pt	0 pt
Nivel de actualización de la respuesta, con respecto a la medicina basada en la evidencia	2 pt	1 pt	0 pt

*Puntajes totales entre 4-6: Respuesta "Buena"; **Puntajes totales entre 2-3: Respuesta Regular"; ***Puntajes totales entre 0-1: Respuesta "Mala".

Resultados

Pregunta Clínica N°1: Diagnóstico Precoz Parálisis Cerebral

Se realizaron 3 niveles de preguntas clínicas a los 5 modelos de IA generativa utilizados: Pregunta clínica simple, pregunta clínica + datos de entrenamiento y pregunta clínica en formato de protocolo. Luego, las respuestas entregadas fueron calificadas utilizando la pauta de apreciación definida. La Tabla 3 resume los resultados encontrados.

Esta pregunta clínica tenía una respuesta muy clara y objetiva en la medicina basada en la evidencia, por lo que, en las respuestas catalogadas como malas, las mayores deficiencias estuvieron en las categorías de calidad técnica y actualización de la respuesta con la medicina basada en la evidencia. ChatGPT4o y GPT Teletón destacaron con las respuestas de mayor calidad, especialmente cuando se agregaron los datos de entrenamiento a la pregunta clínica. Perplexity, a pesar de no obtener buenas calificaciones en sus respuestas, destaca porque, junto con su respuesta, entrega un listado de las referencias que utilizó, lo que agrega mayor transparencia al proceso utilizado. Las Tablas 4 y 5 entregan ejemplos de una respuesta catalogada como “Buena” y una respuesta catalogada como “Mala”, respectivamente.

Llamó la atención la respuesta entregada por Gemini Adv a la segunda pregunta clínica (pregunta + datos de entrenamiento), pues señaló que no disponía de la información suficiente para contestar. Se optó por repetir la misma pregunta y el modelo fue capaz de entregar una respuesta a ésta. Los detalles de este error detectado se incluyen en la Tabla 6.

Con respecto a la confección de una res-

puesta con estructura de protocolo, destaca que todas las herramientas de IA generativa fueron capaces de estructurar un documento siguiendo los lineamientos y estructura entregada, con un lenguaje coherente y de fácil comprensión. Nuevamente, las dificultades de algunas de las respuestas estuvieron, principalmente, en la calidad técnica de la respuesta y actualización de ésta acorde a la medicina basada en la evidencia.

Pregunta Clínica N°2: Prescripción Protésica Deficiencias Congénitas Extremidad Superior

Al igual que en la pregunta anterior, se ejecutaron los 3 niveles de preguntas clínicas ya señalados a los 5 modelos de IA generativa utilizados. La Tabla 7 resume los resultados encontrados.

En esta pregunta clínica, ChatGPT 4o fue el modelo que obtuvo las mejores calificaciones, seguido por Gemini Adv. Llama la atención que Gemini Adv, Claude 3.5 y GPT Teletón registraron un deterioro en la calidad de sus respuestas al agregar los datos de entrenamiento. Esto podría explicarse porque los artículos entregados para apoyar la respuesta, al contrario que en la pregunta anterior, no contenían la respuesta textual a la pregunta clínica, sino que ésta se debía inferir a través del análisis y lectura crítica de los artículos. Por otra parte, al ser una respuesta menos concreta y con mayor margen que la anterior, tampoco se detectaron respuestas completamente incorrectas. La Tabla 8 entrega un ejemplo de una respuesta catalogada como “Buena”.

Con respecto a la solicitud de confección de protocolo, nuevamente destaca que todos los modelos fueron capaces de generar documentos coherentes y con la estructura solicitada

Tabla 3. Resultados de las calificaciones de las respuestas para cada herramienta de IA generativa

Pregunta	Modelos IA Generativa				
	ChatGPT 4o	Gemini Adv	Claude 3.5	Perplexity	GPT Teletón
Pregunta clínica simple	Regular	Malo	Malo	Malo	Regular
Pregunta clínica + datos de entrenamiento	Bueno	Malo	Regular	Malo	Bueno
Pregunta clínica en formato protocolo	Bueno	Regular	Regular	Malo	Bueno

Tabla 4. Respuesta de GPT Teletón a la pregunta clínica + datos de entrenamiento, catalogada como "Buena"

¿Cómo se hace el diagnóstico de Parálisis Cerebral en una persona lactante menor de 5 meses de edad? Para ello considera el artículo Novak I, Morgan C, Adde L, Blackman J, Boyd RN, Brunstrom-Hernandez J, Cioni G, Damiano D, Darrah J, Eliasson AC, De Vries LS. Early, accurate diagnosis and early intervention in cerebral palsy: advances in diagnosis and treatment. JAMA pediatrics. 2017 Sep 1;171(9):897-907.	
Respuesta GPT Teletón	<p>El diagnóstico de parálisis cerebral (PC) en lactantes menores de 5 meses de edad, según el artículo de Iona Novak y colaboradores publicado en 2017 en la revista JAMA Pediatrics, se realiza utilizando una combinación de herramientas estandarizadas junto con la historia clínica para predecir el riesgo de PC de manera temprana y precisa.</p> <p>Herramientas de Diagnóstico para Lactantes Menores de 5 Meses</p> <ol style="list-style-type: none"> 1. **Resonancia Magnética (MRI) a término**: Tiene una sensibilidad del 86%-89% para detectar lesiones cerebrales que podrían llevar a un diagnóstico de PC. 2. **Evaluación Cualitativa de Movimientos Generales de Prechtl (GMs)**: Es la herramienta con mayor sensibilidad (98%) para predecir PC. Se basa en la observación de movimientos espontáneos del lactante, particularmente los movimientos "fidgety" que deben estar presentes entre las 9-20 semanas post término. 3. **Examen Neurológico Infantil de Hammersmith (HINE)**: Con una sensibilidad del 90%, evalúa el tono, los reflejos y la calidad de los movimientos del lactante, proporcionando una puntuación que puede indicar riesgo de PC. <p>### Procedimiento de Diagnóstico</p> <ul style="list-style-type: none"> - **Historia Clínica**: Incluye factores de riesgo preconceptionales, perinatales y postnatales que pueden contribuir al desarrollo de PC, como antecedentes de prematuridad, asfixia neonatal, infecciones y otros factores. - **Evaluación Neurológica y Motora**: La combinación de herramientas como la MRI, GMs y HINE proporciona una evaluación integral y precisa del riesgo de PC. Estas evaluaciones deben ser realizadas por un equipo clínico experimentado para asegurar la exactitud del diagnóstico. - **Intervención Temprana**: Ante la sospecha de PC, incluso si no se puede confirmar con certeza, se recomienda el diagnóstico provisional de "alto riesgo de parálisis cerebral" para que el lactante pueda recibir intervenciones específicas y monitoreo regular, optimizando así la neuroplasticidad y minimizando complicaciones. <p>El uso de estas herramientas y enfoques permite un diagnóstico temprano y preciso de la parálisis cerebral, facilitando la intervención oportuna y mejorando los resultados funcionales y neurológicos del lactante.</p>

y las diferencias de calificación estuvieron, principalmente, en los contenidos técnicos de las respuestas.

El *role-playing* es una estrategia utilizada por expertos para mejorar la calidad de las respuestas entregadas por el modelo de IA. Consiste en entregarle a la herramienta un rol específico a desempeñar como, por ejemplo, una profesión en específico. Para probar el uso de *role-playing* y su impacto en la calidad de las respuestas, se agregó una cuarta pregunta, solicitándole al modelo de IA generativa que se colocara en el rol de un médico especialista en rehabilitación con amplia experiencia en prótesis de extremidad superior. Esto no generó un cambio en la calidad de las respuestas entregadas en ninguno de los modelos de IA generativa utilizados.

Por último, hay que mencionar que, en la

sección de "Bibliografía" de la pregunta clínica con formato de protocolo, el modelo de GPT Teletón entregó un artículo de parálisis cerebral, el cual no tenía ninguna relación con la temática de la pregunta. Este error en las referencias es un claro ejemplo de lo que correspondería a una alucinación de IA y puede afectar la validez de los resultados frente a las preguntas clínicas. La Tabla 9 muestra los detalles de esta alucinación.

Discusión

Este artículo tiene como objetivo evaluar y comparar la capacidad de los modelos de IA generativa para responder preguntas clínicas e intentar dilucidar su aporte en la eventual creación de revisiones sistemáticas, protocolos

Tabla 5. Respuesta de Perplexity a la pregunta clínica simple, catalogada como “Mala”

¿Cómo se hace el diagnóstico de Parálisis Cerebral en una persona lactante menor de 5 meses de edad?

Respuesta Perplexity El diagnóstico de parálisis cerebral (PC) en una persona lactante menor de 5 meses de edad es un proceso complejo que implica una evaluación clínica y diagnóstica exhaustiva. A continuación, se presentan los pasos clave para el diagnóstico de PC en bebés prematuros o lactantes menores de 5 meses:

1. **Examen físico**: El médico realizará un examen físico detallado del bebé, incluyendo la evaluación del tono muscular, la coordinación, el equilibrio y la postura. Esto ayudará a identificar posibles signos de parálisis cerebral, como la debilidad muscular, la espasticidad o la falta de coordinación.
2. **Imágenes cerebrales**: La resonancia magnética nuclear (RMN) es el método de imagen más comúnmente utilizado para diagnosticar la PC en bebés prematuros o lactantes menores de 5 meses. La RMN puede detectar anomalías en el cerebro que puedan causar síntomas de PC.
3. **Análisis de sangre**: El médico puede realizar análisis de sangre para identificar posibles causas de la PC, como anomalías genéticas o infecciones.
4. **Pruebas de funcionalidad nerviosa y muscular**: Estas pruebas pueden incluir pruebas de conducción nerviosa y electromiografía para evaluar la función nerviosa y muscular.
5. **Interrogatorio**: El médico interrogará a los padres sobre el desarrollo del bebé, incluyendo problemas durante el embarazo o el parto, y sobre el progreso del desarrollo del niño. Esta información puede ayudar a identificar la causa de la PC.
6. **Evaluación neurológica**: El médico realizará una evaluación neurológica exhaustiva para determinar si el bebé tiene signos de parálisis cerebral. Esto puede incluir la evaluación de la función motora, la coordinación y el equilibrio.
7. **Derivación a especialistas**: Si el médico sospecha que el bebé tiene PC, puede derivarlo a especialistas como un neurólogo pediátrico, un pediatra del desarrollo o un oftalmólogo, dependiendo de los síntomas específicos.
8. **Seguimiento y monitoreo**: El médico realizará un seguimiento y monitoreo regular del bebé para evaluar el progreso y ajustar el tratamiento según sea necesario.

En resumen, el diagnóstico de PC en una persona lactante menor de 5 meses de edad implica una evaluación clínica y diagnóstica exhaustiva que incluye un examen físico, imágenes cerebrales, análisis de sangre, pruebas de funcionalidad nerviosa y muscular, interrogatorio, evaluación neurológica, derivación a especialistas y seguimiento y monitoreo regular

Citations:

- [1] <https://www.msmanuals.com/es/hogar/salud-infantil/trastornos-neurol%C3%B3gicos-en-ni%C3%B1os/par%C3%A1lisis-cerebral>
- [2] <https://espanol.nichd.nih.gov/salud/temas/cerebral-palsy/informacion/diagnostica>
- [3] <https://www.abclawcenters.com/espanol/signos-de-paralisis-cerebral/>
- [4] <https://www.stanfordchildrens.org/es/topic/default?id=cerebral-palsy-90-P05704>
- [5] http://scielo.sld.cu/scielo.php?pid=S0034-75312007000200007&script=sci_arttext
- [6] <https://www.msmanuals.com/es/professional/pediatr%C3%ADa/trastornos-neurol%C3%B3gicos-infantiles/par%C3%A1lisis-cerebral-pc>
- [7] <https://www.cun.es/enfermedades-tratamientos/enfermedades/paralisis-cerebral-infantil>
- [8] http://ve.scielo.org/scielo.php?pid=S0004-06492013000100008&script=sci_arttext
- [9] <https://www.aeped.es/sites/default/files/documentos/11.pdf>
- [10] <https://espanol.nichd.nih.gov/salud/temas/cerebral-palsy/informacion/signos>
- [11] <https://medlineplus.gov/spanish/cerebralpalsy.html>
- [12] <https://kidshealth.org/es/parents/dyskinetic-cp.html>
- [13] <https://www.mayoclinic.org/es/diseases-conditions/cerebral-palsy/symptoms-causes/syc-20353999>
- [14] <https://espanol.nichd.nih.gov/salud/temas/cerebral-palsy/informacion/causas>
- [15] <https://kidshealth.org/es/parents/ataxic-cp.html>
- [16] <https://espanol.nichd.nih.gov/salud/temas/cerebral-palsy/informacion/riesgo>
- [17] <https://espanol.nichd.nih.gov/salud/temas/cerebral-palsy/informacion/sintomas>
- [18] <https://www.mayoclinic.org/es/diseases-conditions/cerebral-palsy/diagnosis-treatment/drc-20354005>

Tabla 6. Respuestas entregadas por Gemini Adv a la pregunta clínica + datos de entrenamiento

¿Cómo se hace el diagnóstico de Parálisis Cerebral en una persona lactante menor de 5 meses de edad? Para ello considera el artículo Novak I, Morgan C, Adde L, Blackman J, Boyd RN, Brunstrom-Hernandez J, Cioni G, Damiano D, Darrah J, Eliasson AC, De Vries LS. Early, accurate diagnosis and early intervention in cerebral palsy: advances in diagnosis and treatment. JAMA pediatrics. 2017 Sep 1;171(9):897-907	
Respuesta Gemini Adv, primer intento	No dispongo de suficiente información sobre esa persona para responder a tu petición. Soy un modelo de lenguaje extenso, y puedo comunicarme y generar texto en respuesta a una amplia gama de peticiones y preguntas, pero mis conocimientos sobre esta persona son limitados. ¿Hay algo más que pueda hacer para ayudarte con esta solicitud?
Respuesta Gemini Adv, segundo intento	El diagnóstico de parálisis cerebral (PC) en lactantes menores de 5 meses es un desafío, ya que los signos clínicos pueden ser sutiles y variables. El artículo de Iona Novak de 2017 en JAMA Pediatrics destaca la importancia de un enfoque multidisciplinario y el uso de herramientas de evaluación específicas para este grupo de edad.

Tabla 7. Resultados de las calificaciones de las respuestas para cada herramienta de IA generativa

Pregunta	Modelos IA Generativa				
	ChatGPT 4o	Gemini Adv	Claude 3.5	Perplexity	GPT Teletón
Pregunta clínica simple	Bueno	Bueno	Bueno	Regular	Bueno
Pregunta clínica + datos de entrenamiento	Bueno	Regular	Regular	Bueno	Regular
Pregunta clínica en formato protocolo	Bueno	Bueno	Regular	Regular	Regular

o guías clínicas. En la literatura científica, encontramos algunos otros artículos con objetivos similares. Por ejemplo, en abril de 2024, Maniaci y colaboradores, realizaron un estudio en el que generaron un algoritmo en IA para optimizar la creación de protocolos para la toma de decisiones clínicas y determinaron que IA generativa puede jugar un rol relevante en la creación de estos documentos, específicamente en generar las búsquedas sistemáticas de la literatura científica y realizar el análisis de los artículos que seleccionen³⁰. Otro estudio publicado en 2023 por Athaluri y colaboradores, analizó la utilidad de ChatGPT en la creación de protocolos en distintas áreas científicas³¹. Para 50 tópicos científicos distintos, la herramienta fue capaz de generar un documento apropiado con una introducción, objetivos, metodología e implicancias. Sin embargo, fue necesario solicitarle que fuese más específico en cuanto a las referencias entregadas. Por otra parte, en 2023, se publicó otro artículo con una metodología similar al nuestro, donde 33 médicos de 17 especialidades generaron 284 preguntas clínicas para

ChatGPT y las catalogaron según precisión y lo completa de estas, utilizando una pauta de apreciación³². Sus resultados muestran que ChatGPT podría tener un rol prometedor en responder este tipo de preguntas, pero que no está exento de limitaciones y errores, por lo que sus respuestas deben ser cotejadas con fuentes confiables. Por último, mencionar un interesante artículo publicado en 2023 donde se sometió a ChatGPT al mismo examen que realizan los médicos estadounidenses para obtener su licencia para ejercer, denominado *United States Medical Licensing Examination* o USMLE³³. Sus resultados muestran que el desempeño de ChatGPT para responder preguntas clínicas de la evaluación fue superior al 60% y que fue capaz de entregar respuestas lógicas y relacionadas al contexto.

Hasta nuestro conocimiento, este es el primer artículo que intenta comparar de manera objetiva 5 modelos distintos de IA generativa, lo que constituye uno de sus grandes aportes. Sólo encontramos un artículo publicado en 2023 por Dziri y colaboradores, donde, a través de una sola pregunta específica, realizan

Tabla 8: Respuesta de ChatGPT 4o a la pregunta clínica con formato de protocolo, catalogada como "Buena".

Por favor, redacta un protocolo para responder la pregunta clínica de cuándo se indica la primera prótesis en un niño con deficiencia transversal de antebrazo. El documento debe incluir las siguientes secciones: Población Objetivo, Marco Conceptual, Objetivo General, Objetivos específicos, Instrumentos de evaluación, Desarrollo del Procedimiento, Recomendaciones, Bibliografía

Respuesta ChatGPT 4o	<p>Protocolo para Responder la Pregunta Clínica de Cuándo se Indica la Primera Prótesis en un Niño con Deficiencia Transversal de Antebrazo</p> <p>i. Población Objetivo</p> <p>Este protocolo está dirigido a pediatras, cirujanos ortopédicos, terapeutas ocupacionales, fisioterapeutas y otros profesionales de la salud involucrados en el cuidado de niños con deficiencia transversal de antebrazo. La población objetivo incluye a niños desde el nacimiento hasta los 18 años que presentan esta condición.</p> <p>ii. Marco Conceptual</p> <p>La deficiencia transversal de antebrazo es una malformación congénita caracterizada por la ausencia total o parcial del antebrazo, afectando significativamente la funcionalidad y calidad de vida del niño. La intervención temprana con prótesis puede mejorar la autonomía, el desarrollo motor y la integración social del niño. El proceso de indicación de una prótesis debe considerar factores físicos, psicológicos y sociales, así como la edad y el desarrollo motor del niño.</p> <p>iii. Objetivo General</p> <p>Establecer un protocolo basado en evidencia para determinar el momento adecuado para la indicación de la primera prótesis en niños con deficiencia transversal de antebrazo, con el fin de optimizar su desarrollo motor, funcionalidad y calidad de vida.</p> <p>iv. Objetivos Específicos</p> <ol style="list-style-type: none"> 1. Identificar los criterios clínicos y funcionales que determinan la necesidad de una prótesis en niños con deficiencia transversal de antebrazo. 2. Evaluar el impacto de la intervención protésica temprana en el desarrollo motor y funcional del niño. 3. Proporcionar recomendaciones basadas en la evidencia para la selección y adaptación de prótesis en esta población. 4. Desarrollar un plan de seguimiento y evaluación para monitorizar el progreso del niño con prótesis. <p>v. Instrumentos de Evaluación</p> <ol style="list-style-type: none"> 1. **Evaluación Clínica:** <ul style="list-style-type: none"> - Historia clínica detallada. - Examen físico completo del miembro afectado y contralateral. - Evaluación del desarrollo motor y funcional. 2. **Cuestionarios y Escalas:** <ul style="list-style-type: none"> - Pediatric Outcomes Data Collection Instrument (PODCI). - Child Amputee Prosthetics Project Functional Status Inventory (CAPP-FSI). - Vineland Adaptive Behavior Scales. 3. **Evaluación Psicológica:** <ul style="list-style-type: none"> - Entrevistas con el niño y los padres para evaluar el impacto psicológico y social de la deficiencia y la prótesis. - Escalas de desarrollo emocional y conductual. <p>vi. Desarrollo del Procedimiento</p> <ol style="list-style-type: none"> 1. **Detección y Diagnóstico:** <ul style="list-style-type: none"> - Realizar una evaluación inicial detallada del niño con deficiencia transversal de antebrazo. - Diagnosticar la severidad de la deficiencia y la presencia de comorbilidades. 2. **Evaluación Multidisciplinaria:** <ul style="list-style-type: none"> - Formar un equipo multidisciplinario que incluya pediatras, ortopedistas, terapeutas ocupacionales, fisioterapeutas y psicólogos. - Evaluar la funcionalidad actual del niño y sus necesidades específicas. 3. **Determinación de la Indicación Protésica:** <ul style="list-style-type: none"> - Basarse en criterios de desarrollo motor, funcionalidad, y necesidades sociales y psicológicas del niño. - Considerar la edad del niño y su capacidad para adaptarse a la prótesis. 4. **Selección y Adaptación de la Prótesis:** <ul style="list-style-type: none"> - Seleccionar la prótesis adecuada según la edad, nivel de actividad y necesidades específicas del niño. - Realizar un proceso de adaptación y entrenamiento para el uso de la prótesis. 5. **Seguimiento y Evaluación Continua:** <ul style="list-style-type: none"> - Establecer un calendario de seguimiento regular para evaluar el progreso del niño. - Realizar ajustes y modificaciones a la prótesis según sea necesario.
----------------------	---

<p>vii. Recomendaciones</p> <ol style="list-style-type: none"> 1. **Intervención Temprana:** <ul style="list-style-type: none"> - Indicar la primera prótesis tan pronto como el niño muestre signos de estar listo para su uso, generalmente alrededor de los 6-12 meses de edad. 2. **Enfoque Multidisciplinario:** <ul style="list-style-type: none"> - Involucrar a un equipo multidisciplinario en el proceso de evaluación, indicación y seguimiento de la prótesis. 3. **Adaptación Personalizada:** <ul style="list-style-type: none"> - Personalizar la prótesis según las necesidades individuales del niño y realizar ajustes regulares. 4. **Apoyo Psicológico y Social:** <ul style="list-style-type: none"> - Proporcionar apoyo psicológico continuo al niño y a su familia para facilitar la adaptación y el uso de la prótesis. 5. **Educación y Entrenamiento:** <ul style="list-style-type: none"> - Ofrecer programas de entrenamiento para el uso de la prótesis y educación sobre su mantenimiento y cuidado.
<p>viii. Bibliografía</p> <ol style="list-style-type: none"> 1. **Referencias Clásicas y Revisiones:** <ul style="list-style-type: none"> - Smith, D. G., & Michael, J. W. (2004). "Prosthetic Rehabilitation." In O&P Rehabilitation. - Dudkiewicz, I., Gabrielov, R., Seiv-Ner, I., Zelig, G., & Heim, M. (2004). "Evaluation of prosthetic usage in upper limb amputees." <i>Disability and Rehabilitation</i>, 26(1), 60-63. 2. **Estudios Recientes:** <ul style="list-style-type: none"> - James, M. A., & Bagley, A. M. (2011). "The role of upper-limb prostheses in children." <i>Journal of Pediatric Orthopaedics</i>, 31(4), 491-494. - Biddiss, E., & Chau, T. (2007). "Upper limb prosthesis use and abandonment: a survey of the last 25 years." <i>Prosthetics and Orthotics International</i>, 31(3), 236-257. 3. **Guías Clínicas:** <ul style="list-style-type: none"> - American Academy of Pediatrics. (2009). "Clinical report: guidelines for the assessment and management of children with limb deficiencies." <i>Pediatrics</i>, 124(4), e1219-e1228. <p>Este protocolo proporciona un enfoque integral para la evaluación y la indicación de la primera prótesis en niños con deficiencia transversal de antebrazo, basado en la evidencia actual y las mejores prácticas clínicas</p>

Tabla 9. Error bibliográfico entregado por GPT Teletón a la pregunta clínica con formato de protocolo

<p>Por favor, redacta un protocolo para responder la pregunta clínica de cuándo se indica la primera prótesis en un niño con deficiencia transversal de antebrazo. El documento debe incluir las siguientes secciones: Población Objetivo, Marco Conceptual, Objetivo General, Objetivos específicos, Instrumentos de evaluación, Desarrollo del Procedimiento, Recomendaciones, Bibliografía</p>	
<p>Respuesta parcial de la sección "Bibliografía" de GPT Teletón, con el error bibliográfico señalado en negrita</p>	<p>viii. Bibliografía</p> <ol style="list-style-type: none"> 1. Davids JR, Wagner LV, Meyer LC, Blackhurst DW. Prosthetic management of children with unilateral congenital below-elbow deficiency. <i>J Bone Joint Surg Am</i>. 2006;88(6):1294-1300. 2. Le JT, Scott-Wyand PR. Pediatric Limb Differences and Amputations. <i>Phys Med Rehabil Clin N Am</i>. 2015;26(1):95-108. 3. Novak I, Morgan C, Fahey M, et al. Early, accurate diagnosis and early intervention in cerebral palsy: advances in diagnosis and treatment. <i>JAMA Pediatr</i>. 2017;171(9):897-907.

una comparación cualitativa de las respuestas entregadas por ChatPGT y Perplexity³⁴. Los autores concluyen que ambos modelos proveen respuestas de similar calidad, con algunos errores, por lo que deben utilizarse con precaución. Al analizar nuestros resultados, impresiona que ChatGPT podría tener una ventaja comparativa con respecto a los

otros modelos de IA generativa, esto porque la calidad de sus respuestas impresiona ser consistentemente mejor que los otros modelos utilizados. Esto podría atribuirse a que OpenAI lleva más tiempo desarrollando y mejorando su modelo. Sin embargo, las otras herramientas han presentado rápidos avances en el último tiempo, por lo que estos resultados podrían

variar en el futuro.

Con respecto a la calificación de la calidad de las respuestas entregadas por los distintos modelos, podemos inferir que nuestros resultados son similares a los reportados por la literatura científica. Es decir, impresiona que la IA generativa podría jugar un rol importante en la generación de respuestas a preguntas clínicas y creación de documentos como revisiones, guías clínicas o protocolos, pero estos documentos no están exentos de riesgos que deben ser abordados. Dentro de estos, cabe mencionar la falta de actualización de las respuestas a la última evidencia disponible, errores en el contenido técnico, la falta de transparencia en el proceso de búsqueda y alucinaciones de la IA. Por esta razón, el uso de IA generativa en esta tarea debe ser considerada como una herramienta de apoyo que puede aumentar la eficacia y rapidez con que se pueden generar este tipo de documentos, pero no reemplaza el juicio y criterio de los expertos. Es necesario realizar más estudios del tema y de mejor calidad metodológica.

Las principales limitaciones de este estudio radican en que es un estudio descriptivo y cualitativo, con un número pequeño de preguntas y evaluadores, lo que puede afectar la validez externa de esta investigación. Sin embargo, constituye un puntapié inicial para incentivar más investigaciones que intenten medir de manera objetiva el impacto que puede tener una herramienta tan importante como la IA en la forma que se ejerce la medicina y, así lograr que las organizaciones de salud puedan incorporar su uso de manera responsable y segura.

Considerando los resultados obtenidos, nuestra recomendación es utilizar los modelos de IA generativa para responder preguntas clínicas o generar documentos como revisiones, guías o protocolos, pero tomando ciertas precauciones importantes. Dentro de estas, cabe mencionar la incorporaciones de medidas para el cuidado y protección de posible información confidencial de pacientes, velar por los aspectos éticos y de seguridad, realizar un entrenamiento específico de las herramientas de IA con información de alta calidad y en el dominio específico de temas de salud a estudiar, utilizar *prompt* que busquen

incentivar la transparencia en los procesos de búsquedas y diversidad de los resultados, e incluir un monitoreo de expertos en todas las etapas críticas del proceso, como la búsqueda de la evidencia y evaluación de los documentos creados. Específicamente en cuanto al ámbito de rehabilitación, consideramos que el uso de modelos de IA generativa podría ser evaluado como herramienta para apoyar en la creación de documentos de sistematización clínica.

Referencias Bibliográficas

1. Gala D, Makaryus AN. The utility of language models in cardiology: a narrative review of the benefits and concerns of ChatGPT-4. *International Journal of Environmental Research and Public Health*. 2023 Jul 25;20(15):6438.
2. Yim D, Khuntia J, Parameswaran V, Meyers A. Preliminary Evidence of the Use of Generative AI in Health Care Clinical Services: Systematic Narrative Review. *JMIR Medical Informatics*. 2024 Mar 20;12(1):e52073.)
3. Reddy S. Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. *Implementation Science*. 2024 Mar 15;19(1):27.
4. Jung LB, Guder JA, Wiegand TLT, et al. ChatGPT passes German state examination in medicine With picture questions omitted. *Dtsch Arztebl Int*. 2023;120:373-374.
5. Bhayana R, Krishna S, Bleakney RR. Performance of ChatGPT on a radiology board-style examination: insights into current strengths and limitations. *Radiology*. 2023;307:e230582.
6. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digit Health*. 2023;2:e0000198.
7. Sarbay I, Berikol GB, Özturan _ IU. Performance of emergency triage prediction of an open access natural language processing based chatbot application (ChatGPT): A preliminary, scenario based cross-sectional study. *Turk J Emerg Med*. 2023;23:156-161.
8. Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med*. 2023;183: 589-596.
9. Gabriel J, Shafik L, Alanbuki A, et al. The utility of the ChatGPT artificial intelligence tool for patient education and enquiry in robotic radical prostatectomy. *Int Urol*

- Nephrol. 2023;55:2717-2732.
10. Mondal H, Mondal S, Podder I. Using ChatGPT for writing articles for patients' education for dermatological diseases: a pilot study. *Indian Dermatol Online J.* 2023;14:482-486.
 11. Cocci A, Pezzoli M, Lo Re M, et al. Quality of information and appropriateness of ChatGPT outputs for urology patients. *Prostate Cancer Prostatic Dis.* Published online July 29, 2023. <https://doi.org/10.1038/s41391-023-00705-y>.
 12. Berg HT, van Bakel B, van de Wouw L, et al. ChatGPT and generating a differential diagnosis early in an emergency department presentation. *Ann Emerg Med.* 2024;83:83-86.
 13. Barak-Corren Y, Wolf R, Rozenblum R, Creedon JK, Lipsett SC, Lyons TW, Michelson KA, Miller KA, Shapiro DJ, Reis BY, Fine AM. Harnessing the Power of Generative AI for Clinical Summaries: Perspectives From Emergency Physicians. *Annals of Emergency Medicine.* 2024 Mar 12.
 14. Heng JJ, Teo DB, Tan LF. The impact of Chat Generative Pre-trained Transformer (ChatGPT) on medical education. *Postgraduate Medical Journal.* 2023 Oct;99(1176):1125-7.
 15. Günay S, Öztürk A, Özerol H, Yiğit Y, Erenler AK. Comparison of emergency medicine specialist, cardiologist, and chat-GPT in electrocardiography assessment. *The American Journal of Emergency Medicine.* 2024 Jun 1;80:51-60.
 16. Temsah MH, Jamal A, Alhasan K, Aljamaan F, Altamimi I, Malki KH, Temsah A, Ohannessian R, Al-Eyadhy A. Transforming Virtual Healthcare: The Potentials of ChatGPT-4omni in Telemedicine. *Cureus.* 2024 May;16(5).
 17. Sridharan, Kannan, and Reginald P. Sequeira. "Artificial intelligence and medical education: application in classroom instruction and student assessment using a pharmacology & therapeutics case study." *BMC Medical Education* 4, no. 1 (2024): 431.
 18. Wójcik S, Rulkiewicz A, Pruszczyk P, Lisik W, Poboży M, Domienik-Karłowicz J. Beyond ChatGPT: What does GPT-4 add to healthcare? The dawn of a new era. *Cardiology journal.* 2023;30(6):1018-25.
 19. Salvagno M, Taccone FS, Gerli AG. Can artificial intelligence help for scientific writing?. *Critical care.* 2023 Feb 25;27(1):75.
 20. Chen Y, Esmaeilzadeh P. Generative AI in medical practice: in-depth exploration of privacy and security challenges. *Journal of Medical Internet Research.* 2024 Mar 8;26:e53008.
 21. Huang Y, Wu R, He J, Xiang Y. Evaluating ChatGPT-4.0's data analytic proficiency in epidemiological studies: A comparative analysis with SAS, SPSS, and R. *Journal of Global Health.* 2024;14.
 22. Goodman RS, Patrinely Jr JR, Osterman T, Wheless L, Johnson DB. On the cusp: Considering the impact of artificial intelligence language models in healthcare. *Med.* 2023 Mar 10;4(3):139-40.
 23. King DR, Nanda G, Stoddard J, Dempsey A, Hergert S, Shore JH, Torous J. An introduction to generative artificial intelligence in mental health care: considerations and guidance. *Current psychiatry reports.* 2023 Dec;25(12):839-46.
 24. Zybaczynska J, Norris M, Modi S, Brennan J, Jhaveri P, Craig TJ, Al-Shaikhly T. Artificial Intelligence-Generated Scientific Literature: A Critical Appraisal. *The Journal of Allergy and Clinical Immunology: In Practice.* 2024 Jan 1;12(1):106-10.
 25. Garg S, Chauhan A. Chat GPT-4: Potentials, barriers, and future directions for newer medical researchers. *The American Journal of the Medical Sciences.* 2024 Jun 1;367(6):406-8.
 26. Masters K, Benjamin J, Agrawal A, MacNeill H, Pillow MT, Mehta N. Twelve tips on creating and using custom GPTs to enhance health professions education. *Medical Teacher.* 2024 Jun 2;46(6):752-6.
 27. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Frontiers in artificial intelligence.* 2023 May 4;6:1169595.
 28. Athaluri SA, Manthana SV, Kesapragada VK, Yarlagadda V, Dave T, Duddumpudi RT. Exploring the boundaries of reality: investigating the phenomenon of artificial intelligence hallucination in scientific writing through ChatGPT references. *Cureus.* 2023 Apr;15(4).
 29. Májovský M, Černý M, Kasal M, Komarc M, Netuka D. Artificial intelligence can generate fraudulent but authentic-looking scientific medical articles: Pandora's box has been opened. *Journal of medical Internet research.* 2023 May 31;25:e46924.
 30. Maniaci A, Saibene AM, Calvo-Henriquez C, Vaira L, Radulesco T, Michel J, Chiesa-Estomba C, Sowerby L, Lobo Duro D, Mayo-Yanez M, Maza-Solano J. Is generative pre-trained transformer artificial intelligence (Chat-GPT) a reliable tool for guidelines synthesis? A preliminary evaluation for biologic CRSwNP therapy. *European Archives of Oto-Rhino-Laryngology.* 2024 Apr;281(4):2167-73.
 31. Athaluri SA, Manthana SV, Kesapragada VK, Yarlagadda V, Dave T, Duddumpudi RT. Exploring the

- boundaries of reality: investigating the phenomenon of artificial intelligence hallucination in scientific writing through ChatGPT references. *Cureus*. 2023 Apr;15(4).
32. Johnson D, Goodman R, Patrinely J, Stone C, Zimmerman E, Donald R, Chang S, Berkowitz S, Finn A, Jahangir E, Scoville E. Assessing the accuracy and reliability of AI-generated medical responses: an evaluation of the Chat-GPT model. *Research square*. 2023 Feb 28.
 33. Gilson, Aidan, Conrad W. Safranek, Thomas Huang, Vimig Socrates, Ling Chi, Richard Andrew Taylor, and David Chartash. "How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment." *JMIR medical education* 9, no. 1 (2023): e45312.
 34. Dziri C, Fingerhut A. Should we accept systematically the text provided by Chat GPT or Perplexity?. *La Tunisie medicale*. 2023;101(03):321-2.