

ChatGPT-4 en oftalmología: la tercera ola de la inteligencia artificial

ChatGPT-4 in Ophthalmology: the Third Wave of Artificial Intelligence

Andrés Queupumil Rodríguez^{1*} <https://orcid.org/0000-0002-1539-6701>

Diego Ricardo Luna Cedillo² <https://orcid.org/0009-0004-3179-3135>

Saul Díaz Valenzuela¹ <https://orcid.org/0009-0005-0706-0713>

Sol Paulina Gómez Zurita¹ <https://orcid.org/0009-0008-0950-4826>

Miriam Irene Soto Salvador¹ <https://orcid.org/0009-0005-9844-5108>

¹Instituto Cubano de Oftalmología Ramón Pando Ferrer. La Habana, Cuba

²Escuela Consalud ENARM. Ciudad de México, México.

*Autor para la correspondencia: anquero123@gmail.com

RESUMEN

El objetivo de este estudio es explorar el impacto de ChatGPT-4 y otros modelos de lenguaje grande en la oftalmología, destacando su potencial para revolucionar la detección, diagnóstico y educación en este campo. A través de una revisión de literatura reciente y análisis de estudios unimodales y multimodales se evidencia que ChatGPT-4 podría mejorar la capacidad de apoyo diagnóstico en oftalmología, superando herramientas previas y ofreciendo un enfoque accesible para la integración de la inteligencia artificial en la práctica clínica. Los estudios pedagógicos también muestran su eficacia en la educación médica, aunque con limitaciones en comparación con la experiencia humana. Los desafíos éticos y la necesidad de una regulación adecuada son críticos para su implementación efectiva. Este estudio subraya la importancia de la "ingeniería de indicaciones" y la personalización de los modelos de lenguaje grande para usos específicos en

oftalmología, resaltando tanto el potencial transformador como los retos asociados con la adopción de la inteligencia artificial conversacional en este campo.

Palabras clave: modelos de lenguaje grande; procesamiento del lenguaje natural; diagnóstico asistido por IA; educación médica en oftalmología; desafíos éticos de la IA.

ABSTRACT

The aim of this study is to explore the impact of ChatGPT-4 and other large language models in ophthalmology, highlighting their potential to revolutionize detection, diagnosis, and education in this field. Through a recent literature review and analysis of unimodal and multimodal studies, it is evidenced that ChatGPT-4 could improve diagnostic support capabilities in ophthalmology, surpassing previous tools and offering an accessible approach to the integration of artificial intelligence into clinical practice. Pedagogical studies also show its efficacy in medical education, albeit with limitations compared to human expertise. Ethical challenges and the need for adequate regulation are critical to its effective implementation. This study underscores the importance of “cue engineering” and customization of large language models for specific uses in ophthalmology, highlighting both the transformative potential and challenges associated with the adoption of conversational artificial intelligence in this field.

Keywords: large language models; natural language processing; AI-assisted diagnosis; medical education in ophthalmology; ethical challenges of AI.

Recibido: 14/04/2024

Aceptado: 07/10/2024

Introducción

Los Modelos de Lenguaje Grande (LLM: *Large Language Models*) son arquitecturas de inteligencia artificial (IA) que emulan a la corteza prefrontal, permitiéndoles procesar y generar lenguaje de manera similar a los humanos. Estos “cerebros

virtuales” se entrenan con enormes conjuntos de datos a través de lenguajes de programación y formalismos matemáticos, lo que les otorga una vasta capacidad de almacenar y generar conocimiento.^(1,2) No obstante, la complejidad y sofisticación de los LLM ha supuesto una barrera de acceso para el público general. En este punto, el Procesamiento del Lenguaje Natural (NLP: *Natural Language Processing*) surge como un “intérprete” esencial del lenguaje humano, siendo el puente entre los LLM y los usuarios.⁽³⁾ ChatGPT, desarrollado por OpenAI en 2022, es un modelo pionero de NLP en formato de *chatbot*,⁽⁴⁾ un programa informático que simula una conversación humana.

Los LLM y el NLP representan la tercera ola en la evolución de la IA en oftalmología. La primera ola se centra en la detección de imágenes de retina, utilizando técnicas como la tomografía de coherencia óptica y la angiografía con fluoresceína, para el diagnóstico y seguimiento de retinopatía diabética, degeneración macular relacionada con la edad y glaucoma en general.^(5,6) La segunda ola, en la actualidad en desarrollo, se enfoca en el segmento anterior; empleando herramientas como la topografía corneal, la microscopía endotelial y el OCT (*Optical Coherence Tomography*) de segmento anterior para el diagnóstico y manejo de casos de cirugía refractiva y enfermedades como el queratocono o las distrofias corneales.⁽⁷⁾ Esta tercera ola se caracteriza por la capacidad de procesar e interpretar grandes cantidades de información textual y multimodal, complementando y potenciando las capacidades de las olas anteriores.

Para aprovechar el potencial de esta tercera ola es importante entender que la interacción con los LLM requiere un pensamiento estructurado y creativo por parte del usuario. La IA no está diseñada para realizar todas las tareas de forma automática. Es aquí que cobra relevancia la “ingeniería de indicaciones” (*prompt engineering*), un proceso fundamental para maximizar el potencial de los LLM en diversas aplicaciones.⁽⁸⁾ Esta disciplina implica el diseño y refinamiento de las instrucciones (*prompts*) brindadas a los LLM, estructurando su contenido y forma para comunicar efectivamente la intención del usuario y guiar al modelo hacia la generación de respuestas óptimas. Aborda aspectos clave de la acción solicitada

como el contexto, el estilo y tono, objetivo o dar ejemplos, con el fin de potenciar la calidad de las indicaciones y, en consecuencia, obtener interacciones más productivas.⁽⁹⁾

La IA sigue siendo una herramienta especializada con barreras de acceso; por ello, la integración de ChatGPT-4 supone un cambio sustancial al ampliar su uso más allá de los límites académicos y del *software* de alto costo. Apoyando esta perspectiva de democratización, en un análisis de tendencias emergentes, The Medical Futurist posiciona a la oftalmología en el cuadrante de especialidades basadas en datos y en tareas repetitivas,⁽¹⁰⁾ lo que destaca su potencial de ser remodelada a procesos más interactivos, creativos e individualizados al paciente.

El objetivo de esta revisión es mostrar exhaustivamente el impacto de ChatGPT-4 y otros grandes modelos en esta especialidad. A través de un análisis de estudios recientes, tanto unimodales como multimodales, en áreas que van desde el diagnóstico clínico hasta la educación médica, se busca dilucidar su potencial transformador. Además, se exploran los desafíos éticos y las consideraciones de precisión que surgen con la adopción de la IA conversacional, para informar su implementación responsable y efectiva en la atención oftalmológica.

Métodos

Para esta revisión sistemática se realizó una búsqueda en Google Académico utilizando el navegador Chrome, se seleccionaron estudios publicados en los últimos dos años en inglés. Se incluyeron trabajos originales que exploran la aplicación de ChatGPT y otras tecnologías de IA en oftalmología, específicamente estudios observacionales y revisiones que evidencian su impacto en diagnóstico, tratamiento y educación médica. Se excluyeron comentarios, editoriales y documentos no accesibles en su versión completa. La búsqueda se basó en términos clave como “ChatGPT”, “GPT-4”, “inteligencia artificial”, “aprendizaje

profundo”, “procesamiento del lenguaje natural”, “oftalmología” y subespecialidades específicas.

Los artículos potenciales se identificaron y resumieron utilizando la extensión Transcript de Chrome. Se emplearon diferentes instancias de ChatGPT y otros modelos de lenguaje para síntesis y redacción supervisada manualmente. Además, se mencionan diversos GPTs disponibles; y se ejemplifican los más relevantes con capturas de pantalla en un dispositivo móvil. Estas figuras (etiquetadas de la figura 1 a la 5) se referencian a lo largo del texto para ilustrar los conceptos discutidos en las secciones de estudios unimodales y multimodales, tanto clínicos como pedagógicos, proporcionando un contexto práctico para el lector.

En el uso de ChatGPT para asistir en el proceso de investigación, se tomaron medidas para garantizar la integridad académica y el uso ético de la tecnología; incluyendo la verificación cuidadosa de la información generada, la atribución adecuada de las fuentes y el mantenimiento del control editorial humano sobre el contenido final. Una discusión más detallada de las consideraciones éticas se proporciona en la sección de consideraciones éticas y regulatorias.

Desarrollo y tecnología de CHATGPT

En 2023, ChatGPT-4 mejora su rendimiento en tareas cognitivas, incorporando la capacidad de introducir datos multimodales, como imágenes (los modelos GPT [*Generative Pre-trained Transformer*], un tipo de red neuronal artificial especializada en generar y procesar lenguaje mediante preentrenamiento con grandes cantidades de datos, estaban anteriormente limitados únicamente a la entrada de texto).⁽⁴⁾ Otro aspecto nuevo son los modelos GPT personalizados (Custom GPTs), que pueden actuar como expertos altamente cualificados en oftalmología si son entrenados con literatura, guías o casos de estudio relacionados; similar a cómo un médico consulta literatura especializada para mantenerse actualizado. Para facilitar estas consultas y el acceso a recursos externos se utilizan las APIs (*Application Programming*

Interfaces o Interfaces de Programación de Aplicaciones), que funcionan como canales de comunicación especializados que permiten a los GPTs interactuar con otros sistemas y bases de datos, de manera análoga a cómo los especialistas consultan con colegas o acceden a bases de datos médicas para brindar una mejor atención.⁽¹¹⁾ El antecesor natural a este proceso es Isabel Pro, un sistema diagnóstico que utiliza un motor de procesamiento del lenguaje natural que abarca más de 6000 enfermedades. Funciona capturando características demográficas del paciente, así como una casilla de texto para ingresar características clínicas anormales. Resultados recientes evidencian que ChatGPT lo ha superado en la capacidad de identificar el diagnóstico correcto y colocarlo en las primeras posiciones de la lista de diferenciales.⁽¹²⁾

Estudios clínicos unimodales

Para evaluar sistemáticamente estas capacidades iniciales, el enfoque unimodal, que se basa en el análisis de un solo tipo de datos (en este caso, información clínica escrita) es el que tiene más representatividad. Las bases de datos públicas, que son repositorios de información clínica recopilada de casos reales previamente atendidos, son analizadas para las áreas de córnea, glaucoma y neurooftalmología, comparando su eficacia diagnóstica con los expertos de cada especialidad, alcanzando valores del 85 %, 72,7 % y el 82 % en cada caso.^(13,14,15) Rojas⁽¹⁶⁾ adopta un enfoque comparativo en uveítis, utilizando la nomenclatura SUN (*Standardization of Uveitis Nomenclature*) para construir 25 casos prototípicos. ChatGPT obtiene una eficacia del 72 % en el diagnóstico, mostrando una concordancia del 91,6 % con los oftalmólogos en los planes de manejo propuestos.

Analizando el rendimiento por subespecialidades, Delsoz⁽¹³⁾ evalúa en córnea el desempeño de ChatGPT en casos de distintos niveles de complejidad (simple, moderado y muy complejo), logrando una precisión del 100 % en casos simples y moderados, y una ligera disminución en casos muy complejos. El mismo investigador alcanza una eficacia adecuada de ChatGPT en comparación con

especialistas en glaucoma. Aquí la precisión del *chatbot* es comparable a la de residentes *senior* de oftalmología.⁽¹⁴⁾ En el campo de la neurooftalmología *Madadi*⁽¹⁵⁾ selecciona 22 casos de una base de datos en línea, abarcando una amplia gama de enfermedades crónicas y agudas comúnmente atendidas por subespecialistas en neurooftalmología. A pesar de un aumento del 8 % en desempeño al considerar diagnósticos diferenciales, el rendimiento general de ChatGPT es inferior.⁽¹⁵⁾ La figura 1 A-D muestra al GPT llamado Ophthalmology Resident, en el cual se pide a través de un caso clínico los diagnósticos consecutivos y diferenciales. Existen otros más específicos como Retina Science Search Assistant o Strabismus and Binocular Vision Disorder Analysis que se enfocan en sus respectivas subespecialidades (fig. 1 E, F,G, H).

The screenshot displays the chat interface for 'Ophthalmology Resident'. On the left, there is a profile card for 'Ophthalmology Resident' by Kai Phillip Kern, with 200+ conversations. Below this are sections for 'Research & Analysis' (Category), 'Conversation starters' (e.g., 'Discuss the latest treatment for diabetic retinopathy.'), and 'Capabilities' (Browsing, DALL-E Images, Data Analysis). The main chat area shows a user message: 'Señala los diagnósticos oftalmológicos consecutivos y diferenciales sobre la descripción del siguiente caso Clínico: se trata de una mujer Blanca de 67 años con antecedentes de hipotiroidismo. También tiene una neuralgia del quinto par derecho que se trata con gabapentina. Acude por dolor en el ojo derecho de una semana de evolución, Vision doble en forma horizontal, ojo rojo y tinnitus. La exploración no está tomada la agudeza visual, y está ligeramente elevada la presión ocular, tiene vasos en sacacorchos en la conjuntiva y un exotropía de 15° con una limitación de la abducción. El fondo de ojo no revela datos patológicos.' The chatbot's response is: 'La presentación clínica de esta paciente sugiere la posibilidad de múltiples diagnósticos oftalmológicos, tanto consecutivos como diferenciales. Los diagnósticos consecutivos primarios son: 1. Neuropatía del VI par craneal (Nervio Abducens): La presencia de exotropía y la limitación en la abducción sugieren fuertemente una parálisis del VI nervio craneal, que inerva al músculo recto lateral. Esta condición puede causar diplopía horizontal, como se describe en el caso. La elevación de la presión intraocular (PIO) puede ser secundaria a diversos factores, incluido el estrés vascular o el aumento de la presión intracraneal, que también podría afectar al nervio abducens. 2. Síndrome de Tolosa-Hunt: Una entidad inflamatoria estéril del seno cavernoso o de la fisura orbitaria superior que puede presentarse con dolor orbitario, oculomotor (incluida la abducción limitada debido a la afectación del VI par), y a menudo...' To the right of the chatbot's response, there are two columns of text: 'Diagnósticos Consecutivos Primarios' and 'Diagnósticos Diferenciales'. The 'Diagnósticos Diferenciales' section lists: '1. Fístula carótido-cavernosa: Los vasos en sacacorchos en la conjuntiva sugieren un flujo sanguíneo alterado, lo que podría estar relacionado con una fístula carótido-cavernosa. Este diagnóstico se considera dada la presentación de tinnitus, que puede asociarse con el ruido de flujo turbulento característico de una fístula. La presión intraocular elevada...' The interface includes a 'Start Chat' button and a message input field at the bottom.

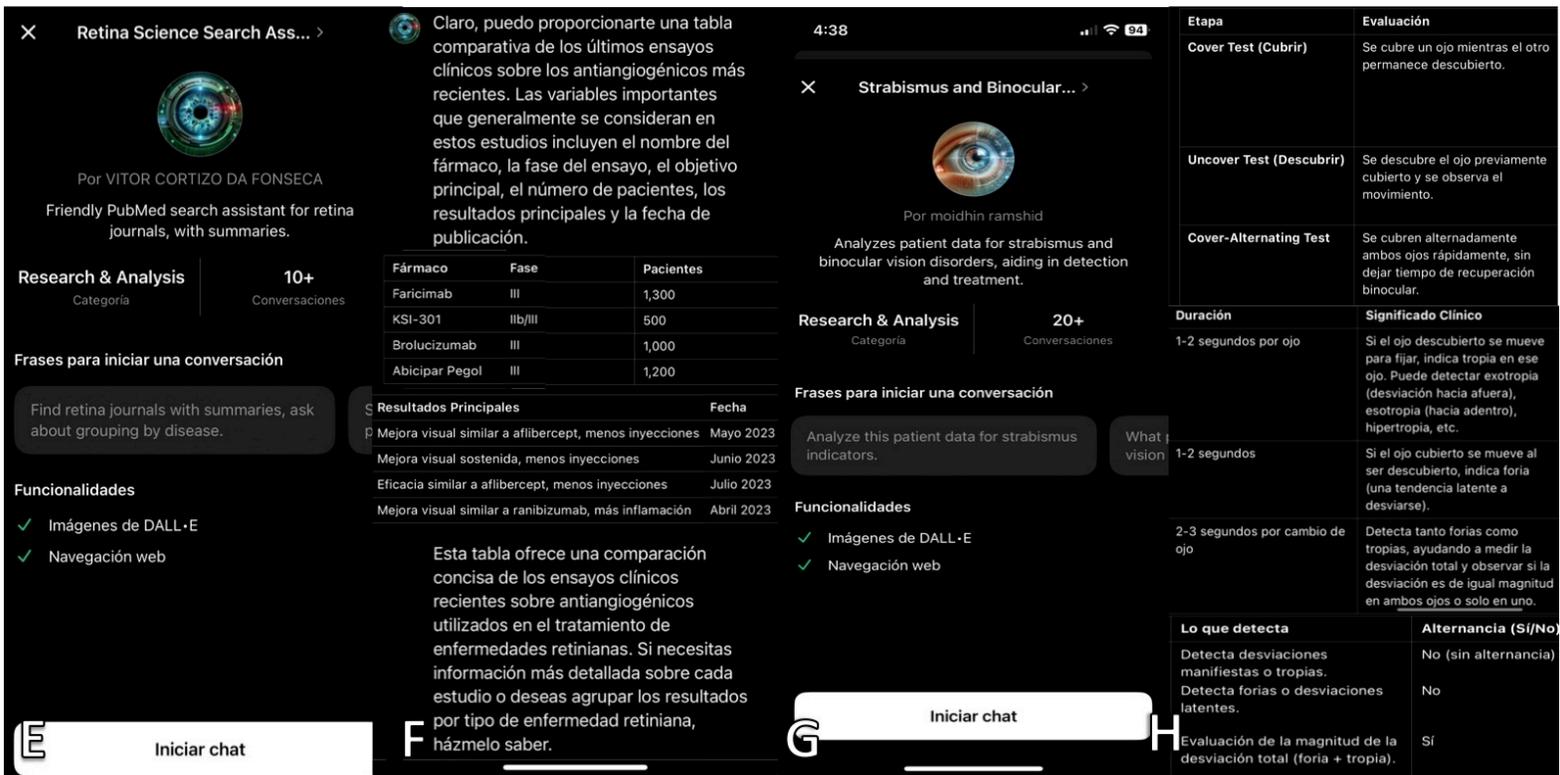


Fig. 1 – A) Interfase del GPT llamado Ophthalmology Resident. B) Planteamiento y descripción del caso clínico con las indicaciones de respuesta. C) Acierta en uno de los diagnósticos, pero no en el principal. D) Pone el diagnóstico real como el primer diferencial. E) Interfase del GPT llamado Retina Science Search Assistant. F) Muestra una tabla comparativa de ensayos clínicos recientes sobre antiangiogénicos en enfermedades retinianas. G) Interfase del GPT llamado Strabismus and Binocular Vision Disorder Analysis. H) Presenta una tabla detallada sobre las pruebas de cover test y sus interpretaciones en el diagnóstico de estrabismo.

En retina, *Raghu*⁽¹⁷⁾ realiza una investigación retrospectiva evaluando diferentes variables continuas (de laboratorio y OCT) con 111 pacientes diabéticos para predecir el riesgo de retinopatía diabética. ChatGPT obtiene una sensibilidad del 67 % y una especificidad del 68 % al incluir valores de espesor subfoveal central, y una sensibilidad del 73 % y especificidad del 54 % sin esta información. A pesar de una excelente fiabilidad en las respuestas (ICC > 0,9), el análisis de kappa de Cohen revela solo una concordancia justa con el diagnóstico clínico. Por otro lado, los rangos de eficacia son inferiores cuando el enfoque se basa en preguntas y respuestas comunes, como en las investigaciones realizadas en pediatría y oculoplastia.^(18,19) *Lim*⁽¹⁸⁾ compara el desempeño de distintos chatbots (ChatGPT-

3.5, 4.0 y Google Bard) en 31 preguntas frecuentes sobre miopía, encontrando porcentajes de respuestas "buenas" del 61,3 %, 80,6 % y del 54,8 %, respectivamente. ChatGPT-4 demuestra una precisión superior en todos los dominios evaluados. En oculoplastia *Ali*⁽¹⁹⁾ evalúa los trastornos del drenaje lagrimal utilizando 21 indicaciones que abarcan aspectos comunes y poco comunes de estos trastornos. Las preguntas cubren afecciones frecuentes como la obstrucción del conducto nasolagrimal adquirida primaria y congénita, así como su causa y manejo. Se concluye que el desempeño de ChatGPT en el mejor de los casos, puede calificarse como promedio, pues solo el 40 % de sus respuestas son completamente correctas, con cierto grado de inexactitud en el 60 % al utilizar una perspectiva más subjetiva. Dejando de lado los resultados diagnósticos y moviéndonos al ámbito terapéutico, *Al-Ashwal*⁽²⁰⁾ revela que ChatGPT tiene un rendimiento destacado en la identificación de interacciones farmacológicas en una serie de 39 casos clínicos de variada complejidad. Si bien proporciona recomendaciones terapéuticas acertadas y elabora planes de manejo general, se observa que la posología no es detallada. La eficacia mostrada subraya la relevancia de una integración precisa de la IA en la práctica clínica. En este contexto, la utilidad de un GPT específico como Pharmacy sugiere un valor añadido significativo para este tipo de evaluaciones farmacológicas.

Estudios pedagógicos unimodales

Los estudios que evalúan el desempeño de ChatGPT en exámenes oftalmológicos basados en preguntas de opción múltiple (MCQ: *multiple-choice questions*) son todos de diseño transversal.^(21,22,23,24) La variable dependiente común en estos estudios es la cantidad de preguntas correctamente respondidas por ChatGPT obtenidas de diversos bancos de preguntas. La mayoría de los estudios utilizan un solo banco, como el *StatPearls* (único estudio que cuantifica los niveles de dificultad del 1-4)^(23,24) o el programa de autoevaluación BCSC de la AAO (*Basic and Clinical Science Course of American Academy of Ophthalmology*)⁽²²⁾ y *Antaki*⁽²¹⁾ emplea dos bancos diferentes: BCSC y *OphthoQuestions*.

Con el objetivo de comprender la evolución de la tecnología, varios estudios comparan el desempeño de diferentes versiones de ChatGPT, como GPT-3,5 y GPT-4.^(22,24) Además, la mayoría de los estudios contrastan el rendimiento de ChatGPT con el de humanos, centrándose principalmente en residentes de oftalmología,^(21,23) mientras que solo *Moshirfar*⁽²⁴⁾ incluye participantes con mayor nivel de experiencia, como *fellows* y especialistas. La figura 2 A-D muestra al GPT Ophthlmo Exam Coach en un ejercicio de preguntas y respuestas de un caso construido de una fuente pública, ejemplificando las herramientas utilizadas en estas evaluaciones.

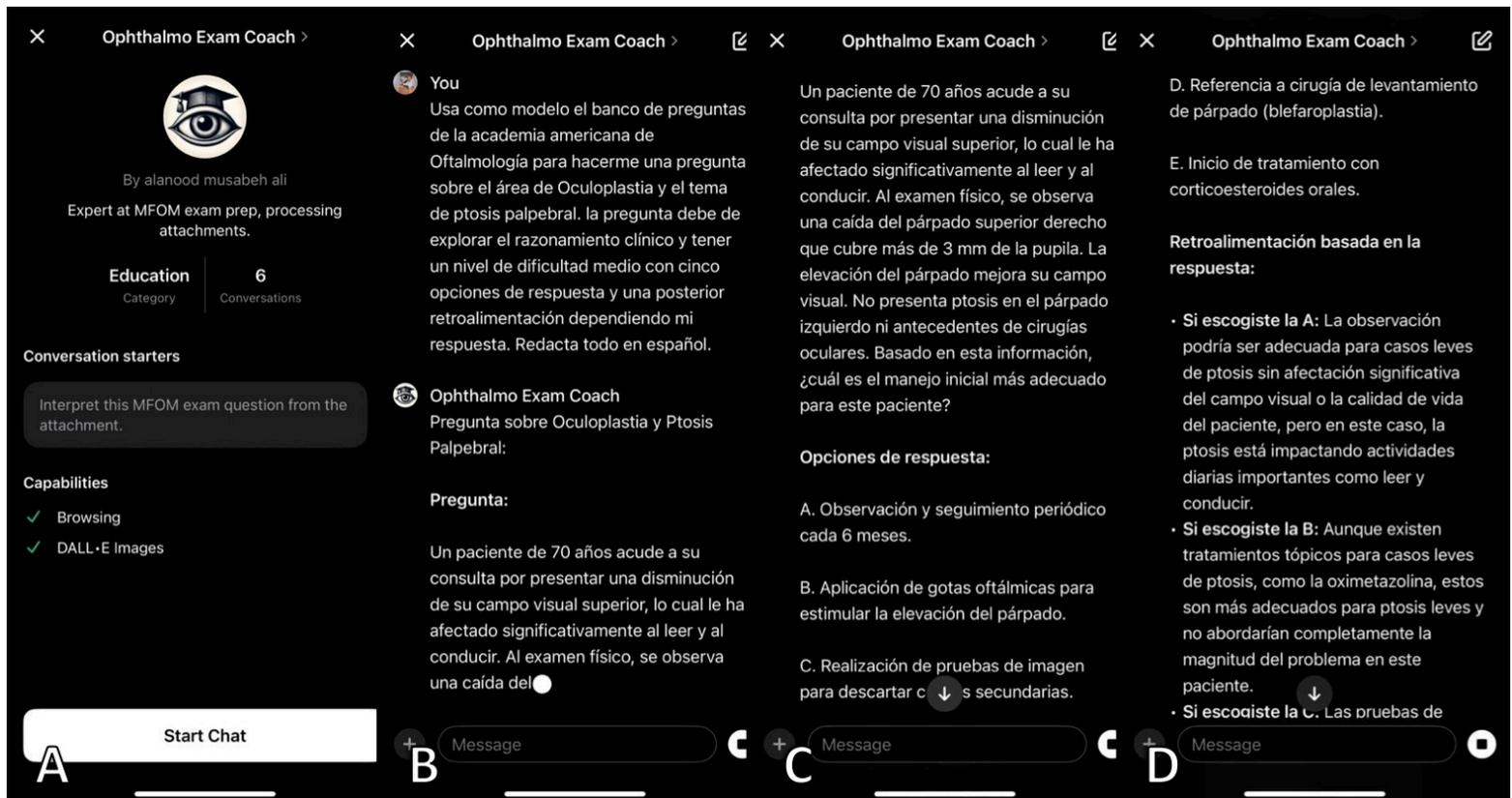


Fig. 2 – A) Interfase del GPT llamado Ophthlmo Exam Coach. B) Indicación de la pregunta conforme a una base de datos específica. C) Planteamiento del caso clínico del área de Oculoplastia con opciones de respuesta D) Parte inicial de la retroalimentación de las respuestas.

Al analizar los resultados específicos de estas evaluaciones comparativas, un hallazgo interesante en el estudio de *Antaki*⁽²¹⁾ es el desempeño de un modelo

llamado Legacy (enero de 2023), el cual refiere ser superior a las versiones de ChatGPT analizadas en el mismo estudio. Sin embargo, en general, el rendimiento de ChatGPT, aún el 4, sigue siendo inferior al de los humanos en la mayoría de los casos.^(21,22,24)

Estudios clínicoterapéuticos multimodales

Es importante señalar que, aunque el análisis multimodal representa un avance prometedor, la capacidad actual de ChatGPT-4 para interpretar imágenes médicas enfrenta limitaciones significativas que requieren consideración. Tras examinar los estudios unimodales, es crucial considerar cómo los avances en modelos multimodales, previamente definidos, están ampliando las capacidades diagnósticas y terapéuticas en la práctica clínica oftalmológica. La incorporación de modelos multimodales representa un avance significativo en el análisis de imágenes médicas. Estos modelos tienen el potencial de revolucionar el diagnóstico, la predicción de resultados y la educación médica al integrar datos de múltiples modalidades. Su desarrollo ha sido impulsado por la aparición de LLMs y modelos de visión expansivos como Segment Anything de Meta, "DALL·E 2" de OpenAI y Stable Diffusion de Stability AI, que exhiben capacidades de texto a imagen para tareas de segmentación o generación.⁽²⁵⁾ La figura 3 A-H muestra al GPT O&AI Ophthalmology Research and Clinical Practice con 2 ejemplos en segmento anterior (imágenes topográficas y tomográficas) y posterior (retinografía de cámara no midriática).

A pesar de estos avances en la integración multimodal ilustrados en la figura 3, la eficacia de ChatGPT-4 en el análisis de imágenes médicas aún no alcanza los niveles esperados. Las evaluaciones iniciales sugieren que la interpretación de imágenes oftalmológicas complejas sigue siendo un área de oportunidad, requiriendo mejoras significativas en el entrenamiento y la capacidad de procesamiento visual. Estas limitaciones se hacen evidentes en diversos campos de la especialidad, como se demuestra a continuación en la evaluación perimétrica del glaucoma y en la interpretación de imágenes de OCT.

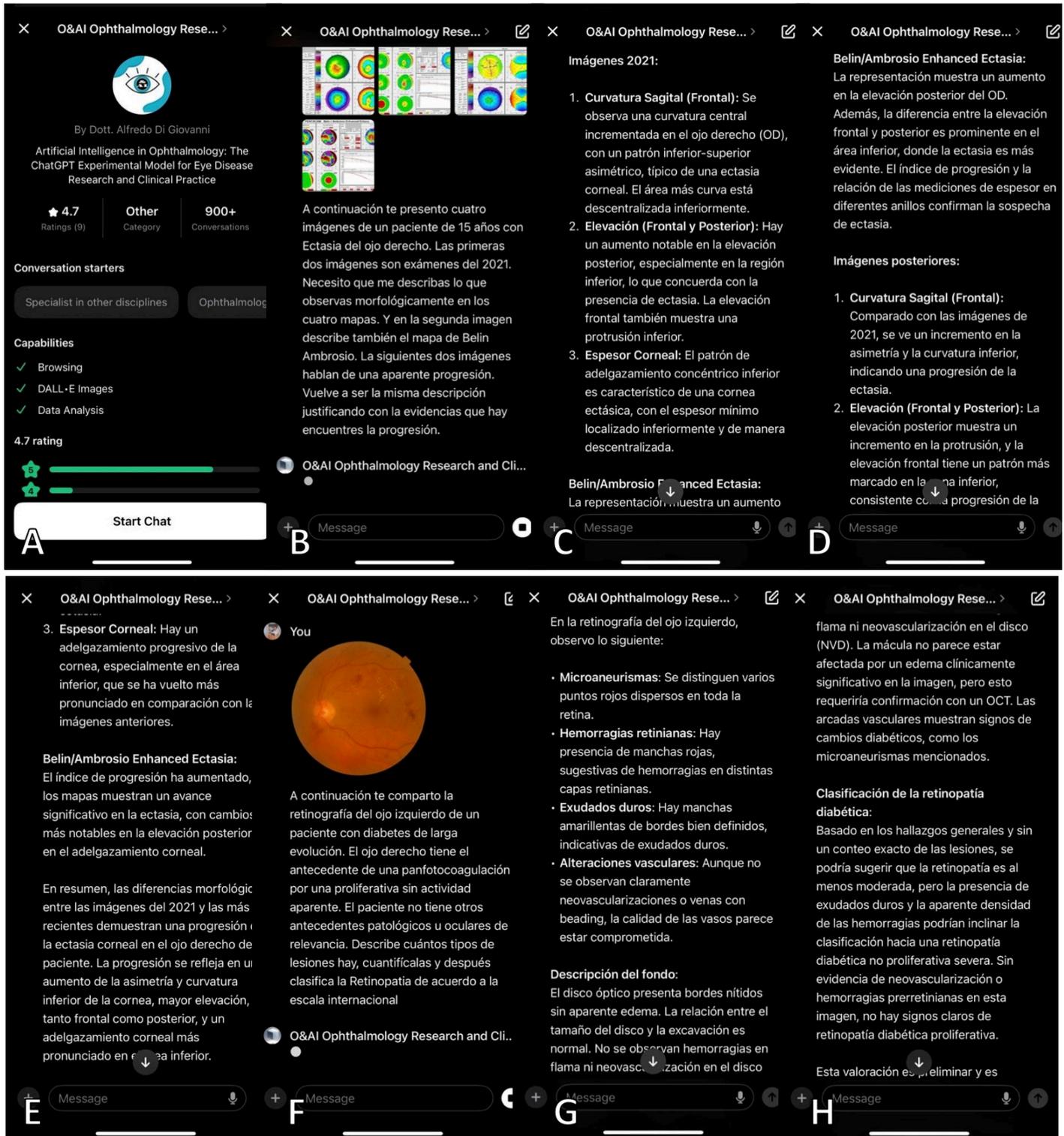


Fig. 3 – A) Interfase del GPT llamado O&AI Ophthalmology Research and Clinical Practice. B) Presentación de ectasia con 4 mapas e indicación de progresión. C) Descripción morfológica del primer mapa. D) corroboración de la ectasia a través del segundo mapa e inicio de descripción de imágenes subsecuentes. E) Redacción final que corrobora la progresión. F) Continuación con caso de

retinopatía diabética con indicación de descripción y estadificación. G) No sigue adecuadamente la indicación (no cuentan las lesiones) pero describe las vistas. H) Clasifica adecuadamente, pero de manera general, este caso de retinopatía diabética.

Un sistema reciente adopta un enfoque multimodal para ayudar a las personas con baja visión, integrando ASR (*Automatic Speech Recognition*), TTS (*Text-to-Speech Synthesis*) y un *bot* de mensajería (Telegram) para crear una interfaz accesible a ChatGPT. Las personas con discapacidad visual pueden interactuar con el sistema a través de mensajes de voz enviados al *bot* de Telegram. Estos mensajes de voz se transcriben a texto utilizando el modelo ASR, y el texto se envía a ChatGPT para obtener una respuesta. La respuesta generada por ChatGPT se convierte en un mensaje de voz utilizando el modelo TTS y se envía de vuelta al usuario a través del *bot* de Telegram. De esta manera, las personas con baja visión pueden aprovechar las capacidades de ChatGPT mediante una interfaz de voz natural, sin necesidad de interactuar directamente con una pantalla o un teclado. El sistema se puede implementar en dispositivos como computadoras o teléfonos inteligentes, brindando una solución accesible y portátil.⁽²⁶⁾

Explorando otras aplicaciones de los sistemas multimodales en el diagnóstico, Huang⁽²⁷⁾ realiza una revisión sistemática de diferentes técnicas de aprendizaje automático (ML: *machine learning*) que permite a las computadoras aprender patrones a partir de datos clínicos mejorando su precisión diagnóstica a medida que procesan más información, y aprendizaje profundo (DL: *deep learning*), un subconjunto del ML que utiliza redes neuronales artificiales con múltiples capas para identificar y clasificar enfermedades oculares complejas, aplicadas a múltiples modalidades de datos, como imágenes de fondo de ojo y campos visuales, para la detección de glaucoma, evaluación de la progresión y estadificación.⁽²⁷⁾ La figura 4 A-D muestra al GPT Advanced Glaucoma Diagnosis Assistant en una evaluación perimétrica.

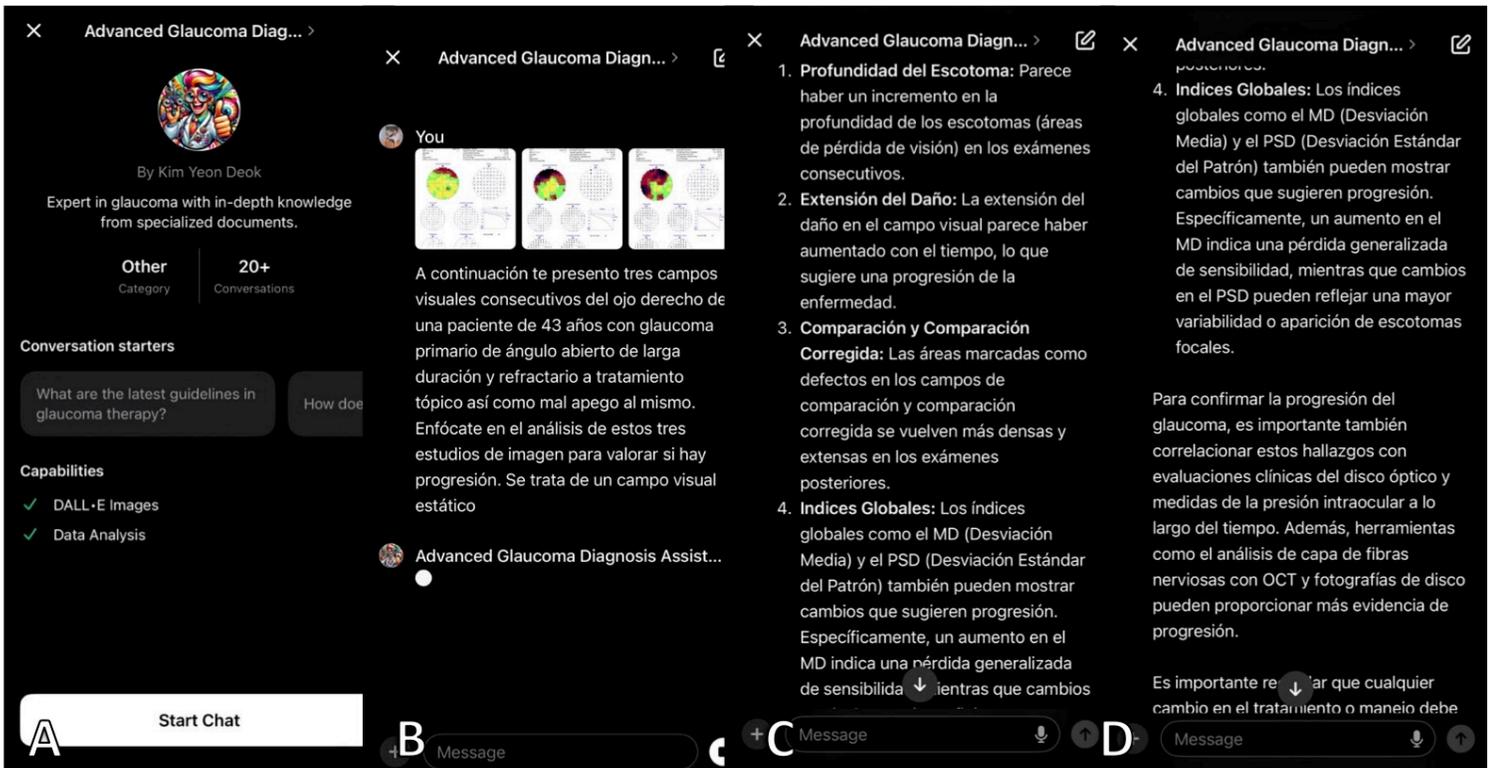


Fig. 4 - A) Interfase del GPT llamado Advanced Glaucoma Diagnosis Assistant. B)

Descripción breve del caso con las imágenes de tres campos visuales estáticos consecutivos del paciente, así como una indicación. C) Corrobora que hay progresión de la enfermedad y lo justifica en cuatro puntos. D) Habla de manera general de los índices y da recomendaciones finales.

La eficacia de precisión diagnóstica en OCT todavía es baja en ChatGPT-4. Aunque estos sistemas de IA tienen el potencial de analizar imágenes de OCT para optimizar la precisión diagnóstica y reducir la variabilidad entre observadores en la detección de enfermedades retinianas, en la actualidad enfrentan desafíos significativos. Para abordar estas limitaciones, Yu⁽²⁸⁾ desarrolla un modelo basado en estas imágenes para el diagnóstico de enfermedades retinianas mediante el reconocimiento de estudios de 80 pacientes que exhibieron características típicas de enfermedades como oclusión de la vena central, edema macular, serosa central o vasculopatía coroidea polipoidea. Se concluye que la forma de presentar la información es crucial para mejorar estos resultados, dado el alto nivel de experiencia requerido para la interpretación. La figura 5 A-D muestra al GPT OCT Master evaluando dos casos.

Existe otro llamado Medical Images Pro Enhanced para imágenes diagnósticas complementarias.

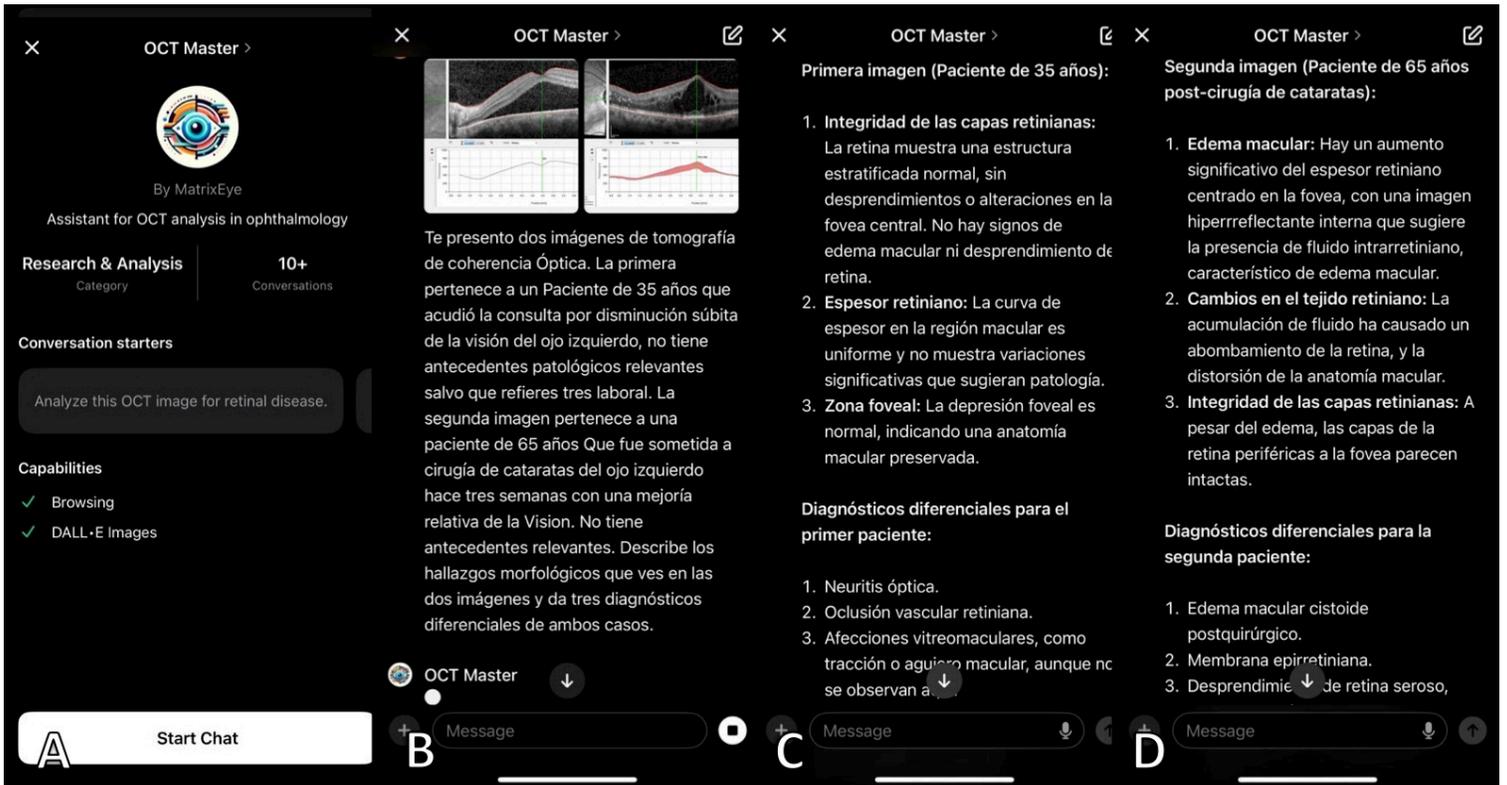


Fig. 5 – A) Interfase del GPT llamado OCT Master. B) Descripción de dos casos clínicos con sus respectivos cortes horizontales de OCT. C) Describe moderadamente la primera imagen, pero no da el diagnóstico correcto en los diferenciales. D) Describe adecuadamente la imagen segunda y acierta en su primer diagnóstico diferencial.

En catarata, la IA tradicionalmente se ha enfocado al cálculo del lente intraocular, hoy el enfoque multimodal hace análisis de los videos quirúrgicos. *Matton*⁽²⁹⁾ crea una gran base de datos llamada BigCat con 3,9 millones de cuadros anotados de 190 videos para identificar instrumentos quirúrgicos y desarrollar sistemas automatizados de evaluación quirúrgica. *Touma*,⁽³⁰⁾ con una red neuronal convolucional (CNN: *Convolutional Neural Network*), un tipo de red neuronal artificial especialmente eficaz en el procesamiento y análisis de imágenes, en 122 videos, logra un área bajo la curva de 0,855, identificando correctamente el 81 % de las 10 fases quirúrgicas propuestas, y una precisión del 96 % cuando clasifica una fase,

mostrando el potencial para automatizar el análisis de videos y mejorar la formación quirúrgica.

Por su parte, *Wang*⁽³¹⁾ propone estimar el tiempo restante en tiempo real, utilizando CNN y memoria a largo-corto plazo (LSTM: *Long Short-Term Memory*), un tipo de red neuronal artificial diseñada para reconocer patrones en secuencias de datos como texto, genomas o series temporales, sin anotaciones manuales. En 2620 videos de tres cirujanos se obtiene un error de 19,4 segundos, superando los métodos estadísticos. Su modelo preentrenado con un cirujano se transfiere eficazmente a otros mediante un ajuste fino con solo el 20 % de datos, lo que es prometedor para estimar el tiempo quirúrgico restante de forma práctica. No obstante, estos avances prometedores y a pesar de reducir en tamaño y tiempo distintos archivos de video, al momento de esta revisión, no hay ningún GPT capaz de procesar uno o varios pasos de la intervención quirúrgica.

Consideraciones éticas y regulatorias

Más allá de los aspectos técnicos y clínicos discutidos, es imperativo abordar los desafíos éticos y regulatorios que surgen con la implementación de ChatGPT en oftalmología. La adopción de esta tecnología va más allá de la simple automatización de procesos; plantea consideraciones complejas sobre integridad académica, protección de datos y equidad en el acceso.

Una de las preocupaciones expresadas por *Hosseini*⁽³²⁾ es la posible distorsión de hallazgos de investigación y la pérdida de credibilidad en la comunidad científica por el uso inadecuado de ChatGPT. Perspectiva real pero contrarrestada con la de "verdaderos expertos", que según *Azaria*⁽³³⁾ son aquellos capaces de identificar y corregir errores; como citas falsas, incapacidad de autocorrección, perpetuación de sesgos, dependencia de información desactualizada, respuestas viciadas o pobre razonamiento lógico; todo a través de escrutinio y supervisión continua.

En contraste con esta aproximación académica rigurosa, otro desafío surge de voces pseudocientíficas, como *influencers* de YouTube, que promueven tácticas narcisistas de "desrobotización" de textos académicos generados por IA.⁽³⁴⁾ Esta

perspectiva es lo que ha llevado a algunas revistas científicas a establecer regulaciones que exigen transparencia para evitar este nuevo neoplagiarismo y obviando que la integridad académica se preserve al no esconder la nueva realidad, sino reportándola y sometiénola a la clásica revisión por pares.

Finalmente, un aspecto crucial a considerar es la protección de datos personales y la potencial regulación de ChatGPT como dispositivo médico, según plantea la guía de la Agencia Reguladora de Productos Sanitarios y Medicamentos del Reino Unido (MHRA: *Medicines and Healthcare products Regulatory Agency*).⁽³⁵⁾ Este marco normativo robusto aseguraría transparencia, interpretabilidad, adaptabilidad y control de riesgos para el paciente. Adicionalmente, las barreras económicas inherentes a las suscripciones mensuales (20-30 USD) podrían limitar su accesibilidad, especialmente en países de bajos ingresos. Ante estos desafíos regulatorios y económicos, es deseable que la comunidad científica asuma un papel proactivo en la regulación de estas tecnologías.

Conclusiones

Se ha destacado la emergencia de los LLM y su aplicación en la oftalmología como la tercera ola de la IA, enfocándose en la capacidad revolucionaria de ChatGPT-4 de procesar e interpretar grandes cantidades de información. Esta revisión exhaustiva indica que ChatGPT-4, con su capacidad para integrar datos multimodales y personalizarse a través de APIs, muestra potencial como herramienta de asistencia en oftalmología, complementando sistemas anteriores y proporcionando apoyo diagnóstico en diversas subespecialidades, aunque aún requiere orientación profesional. Sin embargo, el análisis también revela limitaciones en la interpretación de imágenes médicas complejas y la necesidad de supervisión experta para maximizar su utilidad y precisión.

Los desafíos éticos y las consideraciones sobre la precisión son cruciales, enfatizando la importancia de una implementación responsable para no comprometer la integridad del paciente. En este contexto de oportunidades y

desafíos, esta revisión subraya el potencial transformador de la IA en oftalmología, y llama a una regulación cuidadosa y capacitación continua. Estos hallazgos responden al objetivo inicial de explorar el impacto de ChatGPT-4 en oftalmología, mostrando su potencial como herramienta de apoyo en diagnóstico y educación, al tiempo que enfatizan la necesidad de supervisión experta y desarrollo ético.

Agradecimientos

A las profesoras Ceija Molina (especialidad retina), Mildrey Moreno (especialidad córnea) e Isabel Obret (especialidad glaucoma) por su contribución de imágenes multimodales de sus respectivas especialidades y consultas. Las imágenes se utilizan con la autorización expresa de los especialistas, manteniendo los principios de confidencialidad y anonimización de la información clínica, según las pautas internacionales para el uso de datos en publicaciones científicas.

Referencias bibliográficas

1. Jin K, Yuan L, Wu H, Grzybowski A, Ye J. Exploring large language model for next generation of artificial intelligence in ophthalmology. *Front Med.* 2023;10. DOI: [10.3389/fmed.2023.1291404](https://doi.org/10.3389/fmed.2023.1291404)
2. Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT-Reshaping medical education and clinical management. *Pak J Med Sci.* 2023;39(2):605. DOI: [10.3389/fmed.2023.1291404](https://doi.org/10.3389/fmed.2023.1291404)
3. Wong M, Lim ZW, Pushpanathan K, Cheung CY, Wang YX, Chen D, *et al.* Review of emerging trends and projection of future developments in large language models research in ophthalmology. *Br J Ophthalmol.* 2023; DOI: [10.1136/bjo-2023-324734](https://doi.org/10.1136/bjo-2023-324734)
4. Tan TF, Thirunavukarasu AJ, Campbell JP, Keane PA, Pasquale LR, Abramoff MD, *et al.* Generative artificial intelligence through ChatGPT and other large language

- models in ophthalmology: clinical applications and challenges. *Ophthalmol Sci.* 2023;3(4):100394. DOI: [10.1136/bjo-2023-324734](https://doi.org/10.1136/bjo-2023-324734)
5. Abràmoff MD, Cunningham B, Patel B, Eydelman MB, Leng T, Sakamoto T, *et al.* Foundational considerations for artificial intelligence using ophthalmic images. *Ophthalmology.* 2022;129(2):e14-32. DOI: [10.1016/j.opthta.2021.08.023](https://doi.org/10.1016/j.opthta.2021.08.023)
6. Li Z, Guo X, Zhang J, Liu X, Chang R, He M. Using deep learning models to detect ophthalmic diseases: A comparative study. *Front Med.* 2023;10:1115032. DOI: [10.3389/fmed.2023.1115032](https://doi.org/10.3389/fmed.2023.1115032)
7. Shao Y, Jie Y, Liu ZG, Groups COIS, Intelligent Medicine Branch of Chinese Medicine Education Association. Guidelines for the application of artificial intelligence in the diagnosis of anterior segment diseases (2023). *Int J Ophthalmol.* 2023;16(9):1373. DOI: [10.18240/ijjo.2023.09.03](https://doi.org/10.18240/ijjo.2023.09.03)
8. White J, Fu Q, Hays S, Sandborn M, Olea C, Gilbert H, *et al.* A prompt pattern catalog to enhance prompt engineering with chatgpt. *ArXiv Prepr ArXiv230211382.* 2023; DOI: [10.48550/arXiv.2302.11382](https://doi.org/10.48550/arXiv.2302.11382)
9. Kanika B K. 9 ChatGPT Frameworks To Master Prompt Engineering. *Medium;* 2023 [acceso 05/04/2024]. Disponible en: <https://medium.com/@KanikaBK/9-frameworks-to-master-chatgpt-prompt-engineering-e2fac983bc61>
10. Mesko B. The Medical Futurist. *Future of Medicine. The Impact Of Digital Health Technologies On The Future Of Medical Specialties In One Infographic.* The Medical Futurist; 2023 [acceso 05/04/2024]. Disponible en: <https://medicalfuturist.com/towards-creativity-in-healthcare-the-impact-of-digital-technologies-on-medical-specialties-in-an-infographic/>
11. AI with Tyler. The Benefits of Custom GPT Models Over ChatGPT Plugins [video]. *Youtube;* 2024. (Canal AI with Tyler). Disponible en: <https://www.youtube.com/watch?v=3h5gzNkuyQM>
12. Balas M, Ing EB. Conversational AI models for ophthalmic diagnosis: comparison of ChatGPT and the Isabel Pro differential diagnosis generator. *JFO Open Ophthalmol.* 2023;1:100005. DOI: [10.1016/j.jfop.2023.100005](https://doi.org/10.1016/j.jfop.2023.100005)

13. Delsoz M, Madadi Y, Raja H, Munir WM, Tamm B, Mehravaran S, *et al.* Performance of ChatGPT in diagnosis of corneal eye diseases. *Cornea*. 2024;43(5):664-70. DOI: [10.1097/ICO.0000000000003492](https://doi.org/10.1097/ICO.0000000000003492)
14. Delsoz M, Raja H, Madadi Y, Tang AA, Wirostko BM, Kahook MY, *et al.* The use of ChatGPT to assist in diagnosing glaucoma based on clinical case reports. *Ophthalmol Ther*. 2023;12(6):3121-32. DOI: [10.1007/s40123-023-00805-x](https://doi.org/10.1007/s40123-023-00805-x)
15. Madadi Y, Delsoz M, Lao PA, Fong JW, Hollingsworth T, Kahook MY, *et al.* ChatGPT assisting diagnosis of neuro-ophthalmology diseases based on case reports. *medRxiv*. 2023; DOI: [10.1101/2023.09.13.23295508](https://doi.org/10.1101/2023.09.13.23295508)
16. Rojas-Carabali W, Cifuentes-González C, Wei X, Putera I, Sen A, Thng ZX, *et al.* Evaluating the diagnostic accuracy and management recommendations of ChatGPT in uveitis. *Ocul Immunol Inflamm*. 2023;1-6. DOI: [10.1080/09273948.2023.2253471](https://doi.org/10.1080/09273948.2023.2253471)
17. Raghu K, Tamilselvi S, Devishamani CS, Sucheta M, Rajalakshmi R, Raman R. The utility of ChatGPT in diabetic retinopathy risk assessment: a comparative study with clinical diagnosis. *Clin Ophthalmol*. 2023;4021-31. DOI: [10.2147/OPTH.S435052](https://doi.org/10.2147/OPTH.S435052)
18. Lim ZW, Pushpanathan K, Yew SME, Lai Y, Sun CH, Lam JSH, *et al.* Benchmarking large language models' performances for myopia care: a comparative analysis of ChatGPT-3.5, ChatGPT-4.0, and Google Bard. *EBioMedicine*. 2023;95. DOI: [10.1016/j.ebiom.2023.104770](https://doi.org/10.1016/j.ebiom.2023.104770)
19. Ali MJ. ChatGPT and lacrimal drainage disorders: performance and scope of improvement. *Ophthalm Plast Reconstr Surg*. 2023;39(3):221-5. DOI: [10.1097/IOP.0000000000002418](https://doi.org/10.1097/IOP.0000000000002418)
20. Roosan D, Padua P, Khan R, Khan H, Verzosa C, Wu Y. Effectiveness of ChatGPT in clinical pharmacy and the role of artificial intelligence in medication therapy management. *J Am Pharm Assoc*. 2024;64(2):422-8. DOI: [10.1016/j.japh.2023.11.023](https://doi.org/10.1016/j.japh.2023.11.023)
21. Antaki F, Touma S, Milad D, El-Khoury J, Duval R. Evaluating the performance of ChatGPT in ophthalmology: an analysis of its successes and shortcomings. *Ophthalmol Sci*. 2023;3(4):100324. DOI: [10.1016/j.xops.2023.100324](https://doi.org/10.1016/j.xops.2023.100324)

22. Taloni A, Borselli M, Scarsi V, Rossi C, Coco G, Scorcio V, *et al.* Comparative performance of humans versus GPT-4.0 and GPT-3.5 in the self-assessment program of American Academy of Ophthalmology. *Sci Rep.* 2023;13(1):18562. DOI: [10.1038/s41598-023-45837-2](https://doi.org/10.1038/s41598-023-45837-2)
23. Mihalache A, Popovic MM, Muni RH. Performance of an artificial intelligence chatbot in ophthalmic knowledge assessment. *JAMA Ophthalmol.* 2023;141(6):589-97. DOI: [10.1001/jamaophthalmol.2023.1144](https://doi.org/10.1001/jamaophthalmol.2023.1144)
24. Moshirfar M, Altaf AW, Stoakes IM, Tuttle JJ, Hoopes PC. Artificial intelligence in ophthalmology: a comparative analysis of GPT-3.5, GPT-4, and human expertise in answering StatPearls questions. *Cureus.* 2023;15(6). DOI: [10.7759/cureus.40822](https://doi.org/10.7759/cureus.40822)
25. Li X, Zhang L, Wu Z, Liu Z, Zhao L, Yuan Y, *et al.* Artificial general intelligence for medical imaging. *ArXiv*; 2023. DOI: [10.48550/arXiv.2306.05480](https://doi.org/10.48550/arXiv.2306.05480)
26. Kuzdeuov A, Mukayev O, Nurgaliyev S, Kunbolsyn A, Varol HA. ChatGPT for visually impaired and blind. *En IEEE*; 2024. p. 722-7. DOI: [10.1109/ICAIC60209.2024.10463430](https://doi.org/10.1109/ICAIC60209.2024.10463430)
27. Huang X, Islam MR, Akter S, Ahmed F, Kazami E, Serhan HA, *et al.* Artificial intelligence in glaucoma: opportunities, challenges, and future directions. *Biomed Eng OnLine.* 2023;22(1):126. DOI: [10.1186/s12938-023-01187-8](https://doi.org/10.1186/s12938-023-01187-8)
28. Yu H, Fu Y, Jiang B, Fan P, Shen L, Gao L, *et al.* Applications of GPT-4 for accurate diagnosis of retinal diseases through optical coherence tomography image recognition. *BMC Ophthalmology.* 2023; DOI: [10.21203/rs.3.rs-3644163/v1](https://doi.org/10.21203/rs.3.rs-3644163/v1)
29. Matton N, Qalieh A, Zhang Y, Annadanam A, Thibodeau A, Li T, *et al.* Analysis of cataract surgery instrument identification performance of convolutional and recurrent neural network ensembles leveraging BigCat. *Transl Vis Sci Technol.* 2022;11(4):1-1. DOI: [10.1167/tvst.11.4.1](https://doi.org/10.1167/tvst.11.4.1)
30. Touma S, Antaki F, Duval R. Development of a code-free machine learning model for the classification of cataract surgery phases. *Sci Rep.* 2022;12(1):2398. DOI: [10.1038/s41598-022-06127-5](https://doi.org/10.1038/s41598-022-06127-5)
31. Wang B, Li L, Nakashima Y, Kawasaki R, Nagahara H. Real-time estimation of the remaining surgery duration for cataract surgery using deep convolutional neural

networks and long short-term memory. BMC Med Inform Decis Mak. 2023;23(1):80.

DOI: [10.1186/s12911-023-02160-0](https://doi.org/10.1186/s12911-023-02160-0)

32. Hosseini M, Resnik DB, Holmes K. The ethics of disclosing the use of artificial intelligence tools in writing scholarly manuscripts. Res Ethics. 2023;19(4):449-65.

DOI: [10.1177/17470161231180449](https://doi.org/10.1177/17470161231180449)

33. Azaria A, Azoulay R, Reches S. ChatGPT is a Remarkable Tool—For Experts. Data Intell. 2023;1-49. DOI: [10.1162/dint_a_00235](https://doi.org/10.1162/dint_a_00235)

34. Stapleton A. AI Detection Bypass: Uncovering the Only Method That Works! I Tried Them All! [video]. Youtube; 2023[acceso 5/04/2024]. (Canal Andy Stapleton).

Disponible en: <https://www.youtube.com/watch?v=PI9NqITSCac>

35. Medicines & Healthcare products Regulatory Agency. GOV.UK. 2023 [acceso 05/04/2024]. Software and Artificial Intelligence (AI) as a Medical Device (Guidance). Disponible en: <https://www.gov.uk/government/publications/software-and-artificial-intelligence-ai-as-a-medical-device/software-and-artificial-intelligence-ai-as-a-medical-device>

Conflicto de intereses

Los autores declaran que no existe conflicto de intereses.