

Inteligência Artificial na Educação Médica: Explorando o Potencial do ChatGPT como Ferramenta de Estudo em Oftalmologia

Artificial Intelligence in Medical Education: Exploring ChatGPT's Potential as a Learning Tool in Ophthalmology

 Maria Madeira ¹, Margarida Baptista ¹, Marta Correia ¹, João Romana ¹, Mariana Portela ¹, Ana Cabugueira ¹, Marta Guedes ¹

¹ Serviço de Oftalmologia, Hospital de Egas Moniz (Centro Hospitalar de Lisboa Ocidental) - Lisboa, Portugal

Recebido/Received: 2023-10-15 | Aceite/Accepted: 2024-04-11 | Published online/Publicado online: 2024-07-17 | Published/Publicado: 2024-09-27

© Author(s) (or their employer(s)) and *Oftalmologia* 2024. Re-use permitted under CC BY-NC. No commercial re-use.

© Autor (es) (ou seu (s) empregador (es)) e *Oftalmologia* 2024. Reutilização permitida de acordo com CC BY-NC. Nenhuma reutilização comercial.

DOI: <https://doi.org/10.48560/rspo.33232>

RESUMO

INTRODUÇÃO: A introdução recente de grandes modelos de linguagem (LLMs) baseados na inteligência artificial (IA), sendo o ChatGPT o mais popular, despertou o interesse sobre a sua aplicação em Oftalmologia. Este trabalho propõe-se a explorar o contributo do ChatGPT como auxiliar de estudo no internato de Oftalmologia.

MATERIAL E MÉTODOS: Foi apresentado ao ChatGPT 3.5 (OpenAI, Estados Unidos da América) uma simulação de exame de Oftalmologia, constituído por 260 questões de escolha múltipla, distribuídas pelas 13 áreas de conhecimento do Curso de Ciência Básica e Clínica da Academia Americana de Oftalmologia 2022-2023. Especialistas do Centro Hospitalar de Lisboa Ocidental classificaram a justificação apresentada pelo ChatGPT: 1 (muito fraca), 2 (fraca), 3 (satisfatória), 4 (boa) e 5 (muito boa). Procuramos completar a exatidão do modelo (número de questões certas, na experiência direta E1 e/ou prompted E2) com a precisão (consistência inter-resposta em 3 repetições), bem como estudar a qualidade das justificações, como elementos inovadores.

RESULTADOS: A exatidão do ChatGPT foi de 63,1%, aumentando para 64,2% na experiência guiada. Em ambos, o melhor desempenho registou-se nos fundamentos de oftalmologia (75,0% e 76,7%) e o pior em oculoplástica e órbita (46,7% e 55,0%), óptica e reabilitação e uveíte e inflamação (55,0% e 53,3%; 55,0% e 53,3%). Não se verificou diferença estatisticamente significativa entre o formato da pergunta ou o domínio avaliado e o desempenho do modelo. O ChatGPT acertou as respostas todas as três vezes em 69,3% dos casos, duas vezes em 17,2% e apenas uma vez em 13,5% das situações. Em 94,7% as justificações foram consideradas no mínimo satisfatórias (≥ 3), destas 47,4% atingindo a classificação máxima.

CONCLUSÃO: Os nossos resultados corroboram aqueles descritos na literatura sobre LLMs e bancos de questões oftalmológicas. A natureza probabilística, a ausência de treino específico em Oftalmologia, a incapacidade de assegurar o estado da arte e de processar imagens foram as principais limitações apontadas. Reconhecendo a competência do ChatGPT em fornecer respostas certas sobre temas diferenciados em Oftalmologia, consideramos precoce recomendá-lo como material de estudo apropriado para internos. Contudo, demonstra desde já a exatidão, a precisão e a qualidade científica necessárias para possivelmente se tornar um auxiliar, especialmente se no futuro for treinado especificamente em Oftalmologia.

PALAVRAS-CHAVE: Educação Médica Graduada; Inteligência Artificial; Oftalmologia.

ABSTRACT

INTRODUCTION: The recent introduction of large language models (LLMs) based on artificial intelligence (AI), the most popular of which is ChatGPT, has sparked interest in their application in Ophthalmology. The aim of this study is to investigate the contribution of ChatGPT as a learning tool during Ophthalmology residency.

MATERIAL AND METHODS: ChatGPT 3.5 (OpenAI, United States) was used to simulate an Ophthalmology exam, consisting of 260 multiple-choice questions, distributed among the 13 knowledge areas of the American Academy of Ophthalmology's Basic and Clinical Science Course 2022-2023. Specialists from Centro Hospitalar de Lisboa Ocidental rated the justification provided by ChatGPT: 1 (very poor), 2 (poor), 3 (satisfactory), 4 (good) and 5 (very good). We tried to complement the model's accuracy (number of correct questions in the direct experiment E1 and/or prompted E2) with precision (inter-response consistency over 3 repetitions), and to examine the quality of the justifications, as innovative elements.

RESULTS: ChatGPT accuracy was 63.1%, rising to 64.2% in the prompted test. In both tests, the best performance was recorded in fundamentals of ophthalmology (75.0% and 76.7%) and the worst in oculoplastics and orbit (46.7% and 55.0%), optics and rehabilitation and uveitis and inflammation (55.0% and 53.3%; 55.0% and 53.3%). There was no statistically significant difference between the format of the question or the domain assessed and the model's performance. ChatGPT gave the correct answers all three times in 69.3% of cases, twice in 17.2% and only once in 13.5% of situations. In 94.7% of the cases, the justifications were considered at least acceptable (≥ 3), of which 47.4% achieved the maximum score.

CONCLUSION: Our results confirm those described in the literature on LLMs and ophthalmic question banks. The probabilistic nature, the lack of specific training in Ophthalmology, and the inability to ensure the state of the art and to process images were the main limitations identified. Albeit recognizing ChatGPT's competence to provide the right answers on different topics in Ophthalmology, we believe it is too early to recommend it as suitable learning tool for residents. However, it already demonstrates the accuracy, precision and scientific quality needed to possibly become a tool, especially if specifically trained in Ophthalmology in the future.

KEYWORDS: Artificial Intelligence; Education, Medical, Graduate; Ophthalmology.

INTRODUÇÃO

A inteligência artificial (IA) já é usada em Oftalmologia, com aplicações consolidadas em algoritmos diagnósticos (queratocone), cálculo de potência da lente intraocular (algumas fórmulas de quinta geração) e reconhecimento de imagens (rastreamento de retinopatia diabética), entre outras.¹ Recentemente, foram introduzidos ao público modelos de IA que trabalham especificamente a linguagem, treinados em extensas bases de dados e com elevado poder computacional, designados *large language models* (LLMs). Deles, o mais popular é o ChatGPT, assim baptizado pela sua capacidade conversacional, produzindo texto humanizado a partir da análise transformativa de dados pré-existentes (*generative pre-trained transformer*).²

Os LLMs proporcionam ao utilizador uma conversa fluida e instantânea, características que despertaram o interesse da comunidade médica. O seu potencial para doentes (aconselhamento), sistemas de saúde (simplificação de tarefas administrativas) e profissionais (síntese de informação, processamento de texto, sugestões de investigação e educação médica) é alvo de

múltiplos trabalhos científicos atuais.^{3,4} Relativamente à educação médica, os LLMs são capazes de responder a questões, abertas ou de escolha múltipla, providenciar justificações e gerar resumos, tabelas ou perguntas pedagógicas.²

O desempenho do ChatGPT em testes de avaliação de Oftalmologia foi estudado recentemente na literatura internacional.⁵⁻⁸ Contudo, até à data nenhum destes trabalhos se debruçou sobre a consistência ou justificações que o modelo proporciona, limitando-se a métricas de *performance*.

O presente trabalho propõe-se a explorar o contributo do ChatGPT como auxiliar de estudo no internato de Oftalmologia. Debruçando-nos sobre a sua exatidão, precisão e qualidade científica, pretendemos determinar a utilidade do modelo em complementar o estudo.

MATERIAL E MÉTODOS

CHATGPT

Foi utilizada a versão do Chat GPT 3.5 (OpenAI, Estados Unidos da América), gratuita, de forma a não restringir

a acessibilidade. Encontra-se disponível *online* e atualizada a 3 de Agosto de 2023. O seu treino foi baseado em livros e textos de websites até Setembro de 2021.⁹

BANCO DE QUESTÕES

Foram colocadas 260 questões, adaptadas de uma seleção aleatória daquelas incluídas no Curso de Ciência Básica e Clínica da Academia Americana de Oftalmologia 2022-2023.¹⁰ Esta bibliografia não se encontra disponível online, pelo que não está incluída nos dados nos quais o ChatGPT foi treinado. Tratavam-se de questões de escolha múltipla, igualmente distribuídas pelas 13 áreas de conhecimento ou subespecialidades, nomeadamente Atualização em Medicina Geral, Fundamentos e Princípios de Oftalmologia, Óptica Clínica e Reabilitação Visual, Patologia e Tumores Intraoculares, Neurooftalmologia, Oftalmologia Pediátrica e Estrabismo, Oculoplástica e Órbita, Doença Externa e Córnea, Inflamação Intraocular e Uveítes, Glaucoma, Lente e Catarata, Retina e Vítreo e Cirurgia Refrativa. As questões foram apresentadas ao modelo em inglês. O formato da pergunta foi classificado enquanto simples (S) ou caso clínico (C), e as competências avaliadas distribuídas pelos mecanismos de doença (MD), prevenção (P), diagnóstico (D), gestão do doente (GD) e plano terapêutico (T).

SIMULAÇÃO DE EXAME

Para cada questão, foi apresentada a vinheta da pergunta e as 4 alíneas de resposta (1 correta e 3 erradas), e solicitada uma justificação. A experiência foi repetida 2 vezes: na primeira, a questão foi colocada diretamente (E1), e na segunda, antecedida por uma orientação (prompt) que contextualizava a conversa ao âmbito de um exame de Oftalmologia (E2). A orientação fornecida, traduzida em português, consistia em “Olá ChatGPT, vou apresentar-te algumas questões de escolha múltipla sobre Oftalmologia. Por favor responde da perspectiva de um oftalmologista que se submete a um exame. Seleciona a alínea correta e fornece uma justificação”.

A simulação de exame foi repetida 3 vezes, tendo em conta a natureza probabilística do LLM, e uma nova sessão reiniciada entre cada questão, para limitar o viés de memória.

JUSTIFICAÇÕES

Foi pedido aos especialistas de Oftalmologia do Centro Hospitalar de Lisboa Ocidental que classificassem as justificações apresentadas pelo ChatGPT referentes à sua área de diferenciação. Para isso, foi elaborado um questionário em papel com uma escala de 5 pontos de Likert, que foi preenchido por cada um dos especialistas: 1 (muito fraca), 2 (fraca), 3 (satisfatória), 4 (boa) e 5 (muito boa). Os critérios utilizados consistiram na avaliação do desempenho do ChatGPT em identificar corretamente o conceito-chave subjacente a cada vinheta (o que é) e em explicá-lo corretamente (o que é preciso saber sobre o assunto).

ANÁLISE ESTATÍSTICA

Os dados obtidos foram tabulados em Microsoft Excel[®] 2016. A análise estatística, descritiva e inferencial, foi realizada com IBM SPSS[®] versão 26 2019. Foram calculadas a exatidão (número de respostas certas) e a precisão (consistência inter-respostas) das duas experiências com o modelo (E1 e E2), bem como avaliada a qualidade (classificação) da justificação. Para avaliar os possíveis efeitos da área de conhecimento ou formato/competência da pergunta no desempenho do modelo, foram aplicados os testes *one-way* ANOVA e qui-quadrado, respetivamente, tomados como significativos quando valor-*p* < 0,05.

RESULTADOS

A caracterização das 260 perguntas colocadas ao ChatGPT, no que diz respeito ao seu formato e competência avaliada, encontra-se expressa em frequência absoluta e relativa na [Tabela 1](#). Aproximadamente um terço das perguntas eram casos clínicos (n=84; 32,3%) e mais de metade exigiam compreensão dos MD (n=73; 28,1%) ou capacidade para estabelecer um D (n=89; 34,2%).

A exatidão global (E1 + E2) do ChatGPT foi calculada em 63,7% ± 0,8%, registando-se 63,1% ± 1,4%, na E1, aumentando para 64,2% ± 1,2% na E2. Em ambas as experiências, o melhor desempenho registou-se nos Fundamentos e Princípios de Oftalmologia (E1 – 75,0% ± 5%, e E2 – 76,7% ± 7,6%) e o pior em Oculoplástica e Órbita (E1 – 46,7% ± 2,9% e 55,0% ± 10%), Óptica Clínica e Reabilitação Visual (E1 – 55,0% ± 5% e E2 – 53,3% ± 10,4%) e Inflamação Intraocular e Uveítes (E1 – 55,0% ± 5% e E2 – 53,3% ± 7,6%). Por ordem decrescente, o modelo teve um desempenho intermédio em Lente e Catarata (E1 – 71,7% ± 5,7% e E2 – 73,3% ± 5,7%), Atualização em Medicina Geral (E1 – 68,3% ± 7,6% e E2 – 75,0% ± 8,7%), Neurooftalmologia (E1 – 73,3% ± 5,7% e E2 – 66,7% ± 7,6%), Retina e Vítreo (E1 – 66,7% ± 5,8% e E2 – 73,3% ± 2,9%), Patologia e Tumores Intraoculares (E1 – 66,7% ± 11,5% e E2 – 66,7% ± 10,4%), Cirurgia Refrativa (E1 – 63,3% ± 2,9% e E2 – 65,0% ± 5%), Glaucoma (E1 – 63,3% ± 2,9% e E2 – 60,0% ± 0%), Oftalmologia Pediátrica e Estrabismo (E1 – 63,3% ± 2,9% e E2 – 58,3% ± 10,4%) e Doença Externa e Córnea (E1 – 51,7% ± 2,9% e E2 – 58,3% ± 5,8%) ([Fig. 1](#)). Foi encontrada uma diferença estatisticamente significativa entre os resultados obtidos nas diversas áreas de conhecimento (valor-*p* = 0,0001), excepto para Oftalmologia Pediátrica e Estrabismo, Retina e Vítreo e Cirurgia Refrativa.

Sobre a consistência inter-respostas, o ChatGPT acertou globalmente (E1 + E2) as questões todas as 3 vezes em 69,3% dos casos, 2 vezes em 17,2% e apenas 1 vez em 13,5% das situações. A precisão global (E1 + E2) do modelo (situação em que acertou as questões em todas as repetições) foi calculada 69,3%, objetivando-se 68,1% na E1, subindo para 70,5% na E2 ([Fig. 2](#)).

Não foi encontrada uma diferença estatisticamente significativa entre o formato da pergunta (valor-*p* = 0,213; 0,157) ou a competência avaliada (valor-*p* = 0,224; 0,22) e a exatidão ou precisão do modelo, respetivamente.

Tabela 1. Caracterização das perguntas, por área de conhecimento.

	1	2	3	4	5	6	7	8	9	10	11	12	13	Frequência Absoluta (n)	Frequência Relativa (%)
Formato															
S	11	18	14	16	9	13	8	13	12	14	16	15	17	176	67,7
C	9	2	6	4	11	7	12	7	8	6	4	5	3	84	32,3
Total	20	260	100												
Competência															
MD	4	13	13	12	0	2	5	4	4	3	4	3	6	73	28,1
P	1	0	1	0	0	0	0	2	1	0	3	1	6	15	5,8
D	7	2	2	7	14	12	4	4	7	13	5	10	2	89	34,2
GD	3	2	3	1	5	3	6	4	3	2	6	3	2	43	16,5
T	5	3	1	0	1	3	5	6	5	2	2	3	4	40	15,4
Total	20	260	100												

1 - Atualização em Medicina Geral, 2 - Fundamentos e Princípios de Oftalmologia, 3 - Óptica Clínica e Reabilitação Visual, 4 - Patologia e Tumores Intraoculares, 5 - Neurooftalmologia, 6 - Oftalmologia Pediátrica e Estrabismo, 7 - Oculoplástica e Órbita, 8 - Doença Externa e Córnea, 9 - Inflamação Intraocular e Uveítis, 10 - Glaucoma, 11 - Lente e Catarata, 12 - Retina e Vítreo, 13 - Cirurgia Refrativa; S - simples, C - caso clínico; MD - mecanismos de doença, P - prevenção, D - diagnóstico, GD - gestão do doente, T - plano terapêutico.

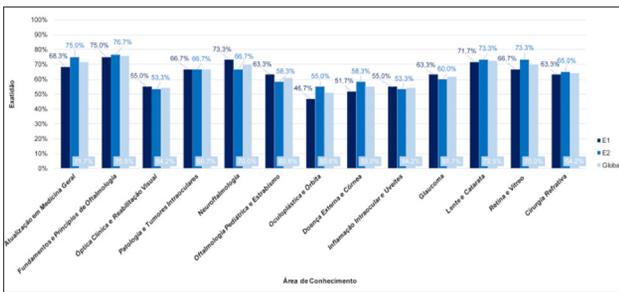


Figura 1. Gráfico de barras da exatidão do ChatGPT na E1 e E2, por área de conhecimento.

E1 - experiência direta, E2 - experiência *prompted*.

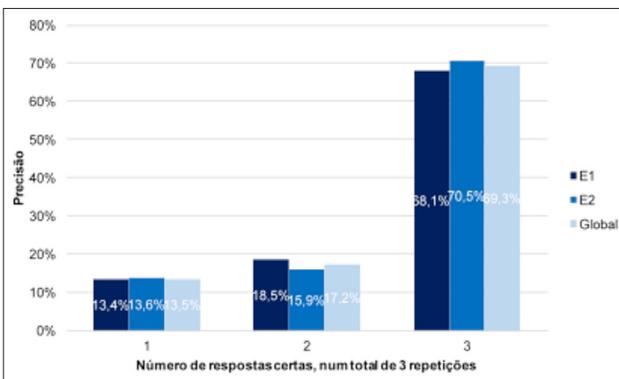


Figura 2. Gráfico de barras da precisão do ChatGPT na E1 e E2.

E1 - experiência direta, E2 - experiência *prompted*.

Relativamente às justificações, 94,7% foram consideradas no mínimo satisfatórias pelos especialistas em Oftalmologia, destas quase metade (47,4%) atingindo a classificação máxima. A qualidade das justificações destacou-se positivamente no Glaucoma, Doença Externa e Córnea e Retina e Vítreo, áreas onde mais de 70% das quais foram marcadas como muito boas, e negativamente na Neurooftalmologia, Fundamentos e Princípios de Oftalmologia,

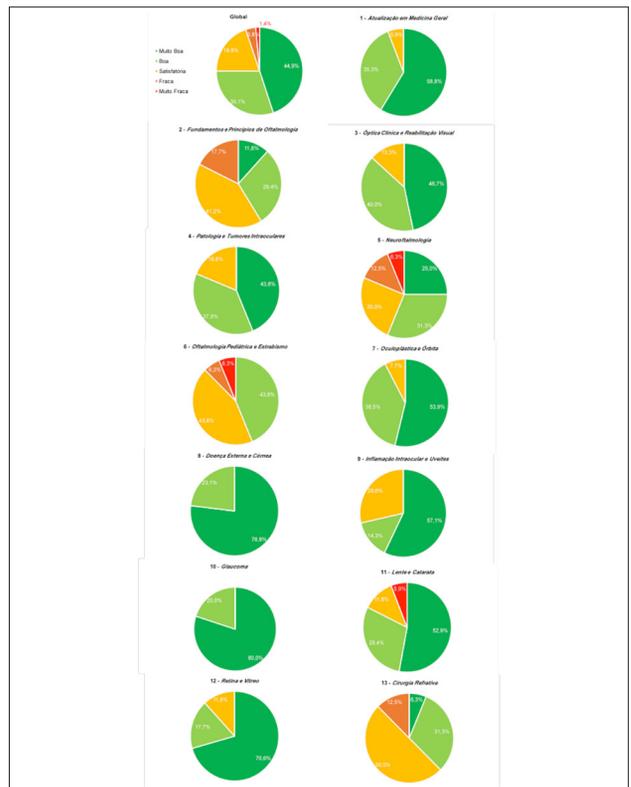


Figura 3. Gráfico circular da classificação da qualidade das justificações apresentadas pelo ChatGPT, por área de conhecimento.

Cirurgia Refrativa, Oftalmologia Pediátrica e Estrabismo, com mais de 10% reportadas fracas ou muito fracas (Fig. 3).

DISCUSSÃO E CONCLUSÃO

Os nossos resultados corroboram aqueles descritos na literatura sobre LLM e bancos de questões oftalmológicas.³⁻⁶ A exatidão global que apuramos (63,7%) para o Cha-

tGPT foi até ligeiramente superior àquelas reportadas nos trabalhos existentes, que a colocam entre os 42,7%-58,8%. Os nossos resultados foram concordantes com a literatura em atribuir uma discreta vantagem à E2, porque é expectável que a utilização de guias (*prompts*) aumente a produtividade do modelo.¹¹ Extravasando a Oftalmologia, a exatidão que calculámos está também de acordo com a descrita noutros trabalhos que se referem a perguntas de Medicina em geral, em que os LLMs são capazes de acertar aproximadamente 60% das questões, apesar da Oftalmologia ser um domínio bastante mais específico.¹²

No nosso estudo, as três melhores áreas de desempenho do ChatGPT foram Fundamentos e Princípios de Oftalmologia, Lente e Catarata e Atualização em Medicina Geral. A maior disponibilidade de recursos online sobre estes temas, por se tratarem de temas transversais ou patologia muito comum, poderá justificar este resultado. A situação inversa ocorre para as três piores áreas de desempenho, Oculoplástica e Órbita, Óptica Clínica e Reabilitação Visual e Inflamação Intraocular e Uveítes. De forma semelhante, a disponibilidade bibliográfica poderá explicar a diferença estatisticamente significativa encontrada no desempenho do ChatGPT entre as áreas de conhecimento.

Contudo, ao nosso conhecimento, nenhum dos trabalhos anteriores se referiu à precisão (consistência inter-resposta), nem à qualidade das justificações apresentadas pelo modelo, elementos que introduzimos no nosso estudo como fator de inovação. Se o nosso objetivo é explorar o potencial do ChatGPT como ferramenta ou assistente de estudo, então consideramos que é preciso garantir que é reprodutível (precisão) e assente no conceito fundamental (*core knowledge*) subjacente à pergunta. Globalmente, a precisão (69,3%) e qualidade das justificações (74,3% boas ou muito boas) que apuramos são satisfatórias para um modelo generalista (por exemplo, na ótica de utilização de um doente ou profissional de saúde não oftalmologista). Contudo, são insuficientes para permitir a utilização do ChatGPT como auxiliar de estudo para um interno de formação específica em Oftalmologia.

Os valores de exatidão, e sobretudo, de qualidade das justificações apresentadas não deixam de surpreender tendo em conta que o modelo não foi treinado em bases de dados de Oftalmologia. Exemplo disso é o nível de conhecimento que o ChatGPT demonstra no domínio cirúrgico. Por outro lado, verificámos alguns casos, apesar de em minoria, em que não identifica ou identifica erroneamente o conceito fundamental subjacente, enrolando-se num ciclo vicioso, de texto supérfluo (Fig. 4). O modelo não tem capacidade de verificação de fontes e pode gerar respostas aparentemente plausíveis, mesmo com informações imprecisas ou desatualizadas, o que pode ser perigoso para o interno. No nosso estudo, as três melhores áreas de justificações do ChatGPT foram Glaucoma, Doença Externa e Córnea e Retina e Vítreo. Verificámos que as áreas onde o ChatGPT respondeu a um maior número de questões corretamente não foram necessariamente aquelas onde se destacou nas justificações. A capacidade pedagógica do modelo pode estar a ser subvalorizada com questões de resposta fechada, mas serão ne-

Durante uma cirurgia de facoemulsificação, o oftalmologista apercebe-se de uma diálise zonular de 90°. Como deve proceder?

- Subir a altura da garrafa e a taxa de fluxo, para manter a profundidade da câmara anterior.
- Converter para cirurgia de extração extracapsular do cristalino.
- Proceder com a facoemulsificação e colocar uma lente intraocular de segmento anterior.
- Colocar um anel de tensão capsular ou retratores de cápsula, completar a facoemulsificação e colocar uma lente intraocular de segmento posterior.

Nesta situação, o cirurgião deve colocar um anel de tensão capsular ou retratores de cápsula, completar a facoemulsificação e inserir uma lente intraocular no segmento posterior. A diálise zonular, que consiste na separação das zónulas que dão suporte ao complexo saco capsular-catarata/lente, cria instabilidade durante a cirurgia de catarata. O cirurgião pode usar um anel de tensão capsular (CTR) ou retratores de cápsula para dar suporte. Uma vez assegurado, pode proceder com a facoemulsificação e inserir uma lente intraocular no saco capsular. **MUITO BOA**

Qual o tipo de defeito campimétrico homônimo produzido por uma lesão na ansa de Meyer?

- Contralateral inferior.
- Contralateral superior.
- Ipsilateral inferior.
- Ipsilateral superior.

A lesão da ansa de Meyer, parte das radiações ópticas do lobo temporal, resulta em hemianopsia hominoma contralateral superior. A ansa de Meyer contém informação visual do campo visual superior do hemisfério contralateral. Por isso, a lesão da ansa de Meyer resulta na perda do campo visual do lado oposto de onde ocorreu a lesão. A lesão da porção inferior das radiações ópticas resulta num defeito campimétrico contralateral inferior. **MUITO FRACA**

Figura 4. Exemplo de interação com o ChatGPT, salientando a dualidade entre duas justificações encontradas.

cessários estudos adicionais para comprová-lo. A qualidade das justificações foi avaliada subjetivamente por especialistas distintos, de áreas variadas e em diferentes momentos da sua diferenciação (desde recém-especialistas a doutorados), o que também pode interferir com os resultados.

A natureza probabilística do modelo e a ausência de treino em bibliografia científica específica são limitações à utilização do ChatGPT como ferramenta de estudo, qualquer que seja a especialidade médica considerada. No que diz respeito à Oftalmologia, a incapacidade de assegurar o estado da arte e de processar imagens são falhas importantes. O modelo torna-se incapaz de acompanhar o desenvolvimento tecnológico crescente que assistiu a esta especialidade ou de reconhecer padrões, característica tão importante no diagnóstico oftalmológico.

A decisão pelo número de questões que avaliámos teve o propósito de se aproximar dos exames escritos internacionais disponíveis para internos de Oftalmologia, não existindo comparador nacional. Tanto o formato do exame americano *Ophthalmic Knowledge Assessment Program* (OKAP), com 260 questões de escolha múltipla de resposta única, como o europeu do European Board of Ophthalmology (EBO), com 52 perguntas seguidas de 5 itens que têm de ser julgados verdadeiros ou falsos, perfazem 260 questões. O nosso estudo avaliou apenas perguntas de escolha múltipla, em que foram sempre fornecidas as opções de resposta ao ChatGPT. No futuro, seria interessante expô-lo também a vinhetas sem opções de respostas ou mesmo testar o seu comportamento com questões abertas ou na criação de materiais de estudo (resumos, tabelas e *flashcards*).

A utilização do inglês para interagir com o ChatGPT teve o objetivo de minimizar a barreira linguística, abrangendo mais dados de treino para expandir a gama de respostas possíveis. Foi demonstrado que a capacidade tradutora do ChatGPT é inferior no domínio biomédico, em relação ao quotidiano conversacional.¹³ Por outro lado, o desafio adicional para internos cuja língua nativa não é a inglesa emerge como limitação adjunta do modelo.

Por fim, a opção pela versão gratuita do ChatGPT 3.5

assentou na garantia da acessibilidade. Contudo, constitui uma limitação, porque já se encontra disponível a versão otimizada ChatGPT 4, embora sujeita a despesa adicional. A capacidade de processar imagens, o treino num conjunto de dados mais vasto (atualizado em Abril de 2023) e o incremento da habilidade em gerar respostas são as principais vantagens do ChatGPT 4. Sendo o objetivo do trabalho investigar uma ferramenta de estudo para um interno de especialidade, cujo objetivo é atingir excelência, seria interessante a utilização da versão mais recente e melhorada do ChatGPT num estudo posterior.

Reconhecendo a competência do ChatGPT em fornecer as respostas certas sobre temas diferenciados em Oftalmologia, consideramos precoce recomendá-lo neste momento como material de estudo apropriado para internos. Contudo, demonstra desde já a exatidão, a precisão e a qualidade científica necessárias para possivelmente se tornar um auxiliar, especialmente no futuro e se for treinado especificamente em Oftalmologia.

CONTRIBUTORSHIP STATEMENT / DECLARAÇÃO DE CONTRIBUIÇÃO:

MM: Revisão bibliográfica, recolha de dados, redação e revisão.

MB, MC e JR: Revisão bibliográfica e recolha de dados.
MP, AC e MG: Redação e revisão.

Todos os autores aprovaram a versão final a ser publicada.

MM: Literature review, data collection, writing and revision.

MB, MC and JR: Literature review and data collection.
MP, AC and MG: Writing and revision.

All authors approved the final version to be published.

RESPONSABILIDADES ÉTICAS

Conflitos de Interesse: Os autores declaram a inexistência de conflitos de interesse na realização do presente trabalho.

Fontes de Financiamento: Não existiram fontes externas de financiamento para a realização deste artigo.

Confidencialidade dos Dados: Os autores declaram ter seguido os protocolos da sua instituição acerca da publicação dos dados de doentes.

Proteção de Pessoas e Animais: Os autores declaram que os procedimentos seguidos estavam de acordo com os regulamentos estabelecidos pelos responsáveis da Comissão de Investigação Clínica e Ética e de acordo com a Declaração de Helsínquia revista em 2013 e da Associação Médica Mundial.

Proveniência e Revisão por Pares: Não comissionado; revisão externa por pares.

ETHICAL DISCLOSURES

Conflicts of Interest: The authors have no conflicts of interest to declare.

Financing Support: This work has not received any contribution, grant or scholarship

Confidentiality of Data: The authors declare that they have followed the protocols of their work center on the publication of data from patients.

Protection of Human and Animal Subjects: The authors declare that the procedures followed were in accordance with the regulations of the relevant clinical research ethics committee and with those of the Code of Ethics of the World Medical Association (Declaration of Helsinki as revised in 2013).

Provenance and Peer Review: Not commissioned; externally peer reviewed.

REFERENCES

- Rosa AM. Inteligência Artificial em Oftalmologia. *Oftalmologia*. 2023;47:75–6. doi:10.48560/rspo.31467.
- Ting DSJ, Tan TF, Ting DSW. ChatGPT in ophthalmology: the dawn of a new era? *Eye* 2024;38:4-7. doi: 10.1038/s41433-023-02619-4.
- Lyons RJ, Arepalli SR, Fromal O, Choi JD, Jain N. Artificial intelligence chatbot performance in triage of ophthalmic conditions. *Can J Ophthalmol*. 2023;S0008-4182(23)00234-X. doi:10.1016/j.cjco.2023.07.016.
- Singh S, Djalilian A, Ali MJ. ChatGPT and ophthalmology: exploring its potential with discharge summaries and operative notes. *Semin Ophthalmol*. 2023;38:503-7. doi:10.1080/08820538.2023.2209166
- 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;3:100324. doi:10.1016/j.xops.2023.100324.
- Raimondi R, Tzoumas N, Salisbury T, Di Simplicio S, Romano MR, North East Trainee Research in Ophthalmology Network (NETRiON). Comparative analysis of large language models in the Royal College of Ophthalmologists fellowship exams. *Eye*. 2023;37:3530-3. doi: 10.1038/s41433-023-02563-3.
- Mihalache A, Popovic MM, Muni RH. Performance of an Artificial Intelligence Chatbot in Ophthalmic Knowledge Assessment. *JAMA Ophthalmol*. 2021;141:589–97. doi: 10.1001/jamaophthalmol.2023.1144.
- Cai LZ, Shaheen A, Jin A, Fukui R, Yi JS, Yannuzzi N et al. Performance of Generative Large Language Models on Ophthalmology Board-Style Questions. *Am J Ophthalmol*. 2023;254:141–9. Doi:10.1016/j.ajo.2023.05.024.
- Chat GPT [accessed from June to August 2023]. Available at: <https://chat.openai.com/>.
- American Academy of Ophthalmology. 2022-2023 Basic and Clinical Science Course, Complete Print Set. San Francisco: AAO; 2022.
- J White, Q Fu, S Hays, M Sandborn, C Olea, H Gilbert, et al. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. *ArXiv*. 2023; Preprint at arXiv:2302.11382v1. doi:10.48550/arXiv.2302.11382.
- Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Ele-

paño C, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023;2:e0000198. doi:10.1371/journal.pdig.0000198.

13. Jiao W, Wang W, Huang J, Wang X, Tu Z. Is ChatGPT a good translator? ArXiv. 2023; Preprint at arXiv:2301.08745v3. doi:10.48550/arXiv.2301.08745.

**Corresponding Author/
Autor Correspondente:**

Maria Madeira

Serviço de Oftalmologia
Hospital de Egas Moniz,
R. da Junqueira 126,
1349-019 Lisboa, Portugal
E-mail: filipamadeira10@hotmail.com



ORCID: 0009-0009-2756-9145