

CARTA - INVESTIGAÇÃO

ChatGPT: Performance da inteligência artificial no exame de obtenção do título de especialista em dermatologia ☆☆☆



Prezado Editor,

O ChatGPT (*Chat Generative Pre-Trained Transformer*), lançado pela OpenAI em 2022, é um modelo avançado de linguagem em inteligência artificial (IA) capaz de interagir em conversas e fornecer respostas originais. Apesar da grande especulação, sua utilização para auxílio na tomada de decisão clínica ainda não é recomendada em virtude da falta de informações sobre seu conhecimento técnico em Medicina e dos potenciais impactos éticos envolvidos.^{1,2}

Como primeiro passo para obter essas informações, o ChatGPT tem sido recentemente testado em provas de título de diferentes especialidades médicas.³⁻⁵ No Brasil, poucos estudos foram realizados nesse sentido e, até o momento, nenhum avaliou seu desempenho na prova de Título de Especialista em Dermatologia (TED).

O TED é obtido por exame anual oferecido pela Sociedade Brasileira de Dermatologia (SBD), tendo como primeira fase eliminatória uma prova teórica composta por 80 questões objetivas, com quatro alternativas cada e apenas uma opção correta. São eliminados do concurso os candidatos que acertarem menos de 60% do total (48 questões corretas).⁶ Realizamos este trabalho com o objetivo de avaliar o desempenho dessa IA nesse exame e ponderar sua utilização na Dermatologia.

Avaliamos a performance do ChatGPT (versão 4) por meio da resolução das questões do TED 2022 e 2023, disponíveis no site da SBD. Foram desconsideradas as questões com imagens, pela ausência dessa funcionalidade, e as anuladas. As questões foram classificadas em dermatologia clínica, laboratorial, cirúrgica e cosmiatria. O aplicativo foi instruído a responder questões de múltipla escolha. Em seguida,

Tabela 1 Distribuição percentual dos acertos por área e ano, com o total de questões entre parênteses

Área	Total	2022	2023
Cirurgia	50 (4)	0 (1)	66,67 (3)
Clínica	77,87 (122)	86,67 (60)	69,35 (62)
Cosmiatria	80 (5)	50 (2)	100 (3)
Laboratorial	60 (15)	44,44 (9)	83,33 (6)
Total	75,34 (146)	79,17 (72)	71,62 (74)

apresentou-se uma questão por vez para que ele escolhesse a alternativa mais adequada. A acurácia geral foi medida como a razão entre o número de respostas corretas e o total de questões. Também foi avaliada a acurácia por ano de prova e para cada área, comparando a média das provas usando o teste *t* de Student ou Mann-Whitney, dependendo da normalidade dos dados, a qual foi verificada por meio do teste de Shapiro-Wilk. Em seguida, avaliamos a variação do desempenho ao longo da prova agrupando todas as questões em grupos de 10 em 10, seguindo a ordem da prova. Em cada grupo calculou-se a porcentagem de acerto, desconsiderando questões anuladas e com imagens daquele grupo. Utilizou-se o teste de Spearman para avaliar a correlação entre a porcentagem de acerto dos grupos e o número da questão inicial do respectivo bloco, assim como uma linha de tendência gráfica utilizando a regressão linear, buscando identificar se houve melhora do desempenho no decorrer da prova.

Foram analisadas um total de 146 questões, das quais 74 do TED 2023 e 72 do TED 2022. Foram excluídas nove questões anuladas e cinco por uso de imagem. A acurácia geral foi de 75,34%, e a média de desempenho na prova de 2022 foi superior à de 2023 ([fig. 1](#)). No teste estatístico de comparação entre as provas, como não houve normalidade dos dados, foi feito o teste de Mann-Whitney, valor de *p* de 0,29. A área com mais acertos foi a de cosmiatria, e a com menos, de cirurgia ([tabela 1](#)). As linhas de tendência mostram inclinação positiva, sugerindo melhora do desempenho no decorrer da prova em ambos os anos ([fig. 2](#)). A análise de correlação de Spearman revelou correlação positiva e moderada, porém não significativa, para o TED 2022 (correlação = 0,43, *p* = 0,28) e TED 2023 (correlação = 0,32, *p* = 0,43).

O desempenho do ChatGPT nos exames de obtenção do TED de 2022 e 2023 foi acima do exigido para a obtenção da aprovação na primeira fase do concurso. No entanto, uma avaliação comparativa dos desempenhos anuais não revelou diferença estatisticamente relevante, impedindo

DOI referente ao artigo:

<https://doi.org/10.1016/j.abd.2023.08.005>

☆ Como citar este artigo: Jabour TBF, Ribeiro Junior JP, Fernandes AC, Honorato CMA, Queiroz MCAP. ChatGPT: Performance of artificial intelligence in the dermatology specialty certificate examination. *An Bras Dermatol*. 2024;99:277-79.

☆☆ Trabalho realizado no Departamento de Dermatologia, Hospital Universitário Onofre Lopes, Universidade Federal do Rio Grande do Norte, Natal, RN, Brasil.

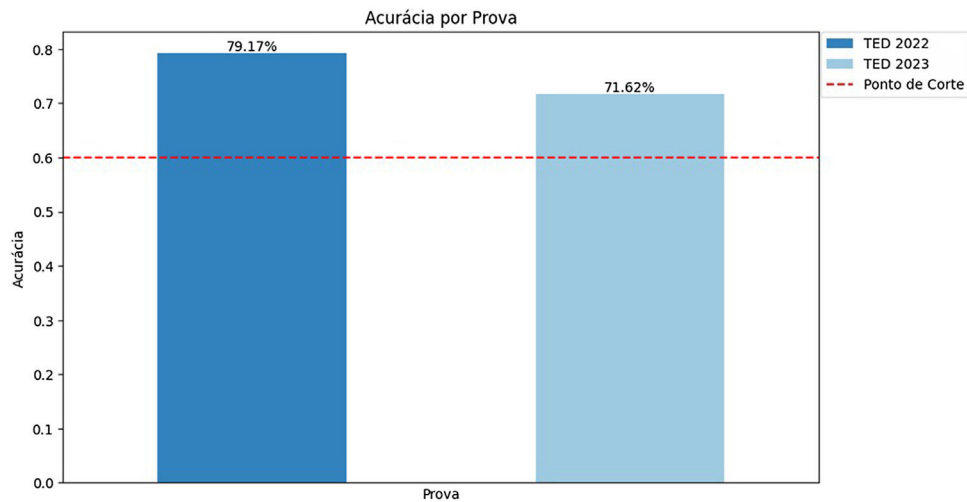


Figura 1 Acurácia do ChatGPT por prova em comparação com o ponto de corte para aprovação no exame.

Porcentagem de acertos por grupo de questões no TED 2022 e 2023

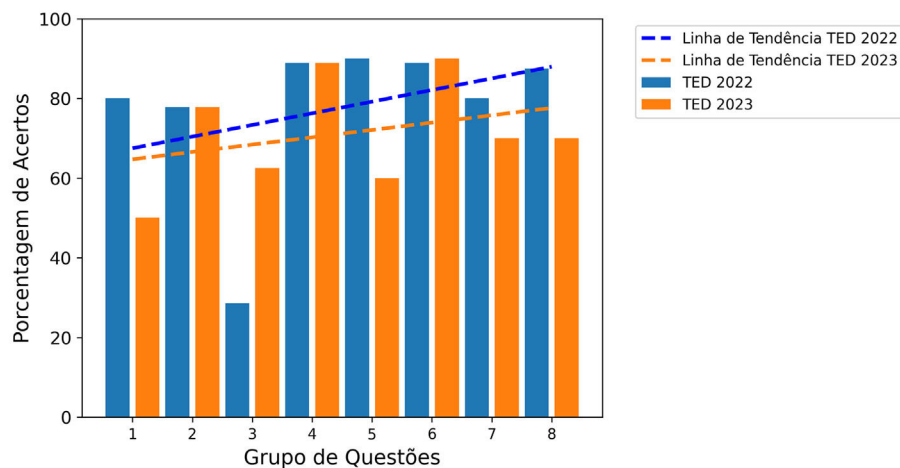


Figura 2 Porcentagem de acertos por grupos de questões ao longo da prova no TED 2022 e 2023, sugerindo melhora do desempenho à medida que a ferramenta foi utilizada.

uma afirmação conclusiva de superioridade de um ano em relação ao outro. A análise de desempenho no decorrer da prova sugere uma tendência à probabilidade de acerto das questões aumentar conforme mais questões são realizadas, embora mais questões sejam necessárias para alcançar valor estatístico mais significativo. As categorias com melhor desempenho foram cosmiatria e clínica, sugerindo predominância dessas áreas no banco de dados de treinamento do ChatGPT.

Os resultados foram consonantes com outros experimentos em avaliações médicas já realizados. Um artigo recém-publicado demonstrou que o ChatGPT tem desempenho equivalente ao de um estudante de Medicina do terceiro ano ao responder às perguntas dos exames de licenciamento médico dos Estados Unidos (USMLE).³ Na Dermatologia, o aplicativo já foi utilizado para resolver questões da prova de certificação de especialidade no Reino Unido. No estudo, o ChatGPT-3.5 obteve pontuação geral de 63,1%, e o ChatGPT-4 obteve 90,5%. A nota de aprovação esperada para o SCE de Dermatologia é de 70%–72%.⁴

Embora os resultados sejam promissores, ainda não há estudos que embasem o uso do ChatGPT na prática médica.² Uma grande limitação para sua aplicação na Dermatologia é a não avaliação de imagens, considerando o caráter visual da especialidade. É preciso cautela na incorporação da IA na prática dermatológica, além de ser essencial a realização de novas pesquisas com colaboração de especialistas em dermatologia e em IA para aprimorar a compreensão dos seus potenciais riscos e benefícios. O exame clínico minucioso e uma boa relação médico-paciente permanecem sendo, até o momento, as melhores ferramentas para a prática clínica segura, confiável e efetiva da especialidade.

Suporte financeiro

Nenhum.

Contribuição dos autores

Thaís Barros Felipe Jabour: Concepção e desenho do estudo; levantamento, análise e interpretação dos dados; revisão crítica da literatura; redação do artigo ou revisão crítica do conteúdo intelectual importante.

José Paulo Ribeiro Júnior: Levantamento, análise e interpretação dos dados; revisão crítica da literatura; redação do artigo ou revisão crítica do conteúdo intelectual importante.

Alexandre Chaves Fernandes: Concepção e desenho do estudo; análise estatística; redação do artigo ou revisão crítica do conteúdo intelectual importante.

Cecília Mirelle Almeida Honorato: Revisão crítica da literatura; redação do artigo ou revisão crítica do conteúdo intelectual importante.

Maria do Carmo Araújo Palmeira Queiroz: Revisão crítica do conteúdo; aprovação final da versão final do manuscrito.

Conflito de interesses

Nenhum.

Referências

1. Ferreira AL, Lipoff JB. The complex ethics of applying ChatGPT and language model artificial intelligence in dermatology. *J Am Acad Dermatol*. 2023;89:e157-8.
2. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare (Basel)*. 2023;11:887.
3. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digital Health*. 2023;2:e0000198.

4. Passby L, Jenko N, Wernham A. Performance of ChatGPT on dermatology specialty certificate examination multiple choice questions. *Clin Exp Dermatol*. 2023;11:197.
5. Mihalache A, Popovic MM, Muni RH. Performance of an artificial intelligence chatbot in ophthalmic knowledge assessment. *JAMA Ophthalmol*. 2023;141:589-97.
6. Sociedade Brasileira de Dermatologia - Edital do 57(Exame para Obtenção do Título de Especialista em Dermatologia (2023) [Internet]. Disponível em: <<https://www.sbd.org.br/wp-content/uploads/2023/04/TED2023edital.pdf>>.

Thaís Barros Felipe Jabour ^a,
José Paulo Ribeiro Júnior ^{b,*},
Alexandre Chaves Fernandes ^c,
Cecília Mirelle Almeida Honorato ^d
e Maria do Carmo Araújo Palmeira Queiroz ^b

^a Centro de Pesquisas Clínicas de Natal, Natal, RN, Brasil

^b Departamento de Dermatologia, Hospital Universitário Onofre Lopes, Universidade Federal do Rio Grande do Norte, Natal, RN, Brasil

^c Instituto Internacional de Neurociências Edmond e Lily Safra, Macaíba, RN, Brasil

^d Departamento de Dermatologia, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, SP, Brasil

* Autor para correspondência.

E-mail: josepaulojr@outlook.com (J.P. Ribeiro Júnior).

Recebido em 29 de junho de 2023; aceito em 11 de agosto de 2023

<https://doi.org/10.1016/j.abdp.2023.11.015>
2666-2752/ © 2023 Publicado por Elsevier España, S.L.U. em nome de Sociedade Brasileira de Dermatologia. Este é um artigo Open Access sob uma licença CC BY (<http://creativecommons.org/licenses/by/4.0/>).

Tinea capitis: tratamento em população pediátrica. Observações em 99 casos ☆☆☆



Prezado Editor,

Tinha do couro cabeludo, ou *Tinea capitis* (TC), é dermatofitose que acomete tanto o couro cabeludo quanto a haste capilar.^{1,2} Os principais agentes causais descritos das TC são dos gêneros *Microsporum* e *Tricophyton*, e a frequência de cada patógeno varia de acordo com a localização geográfica,

fatores ambientais e culturais de cada região e período estudado.³

A apresentação clínica da TC depende da interação entre o organismo causal e a resposta clínica do hospedeiro, resultando em quadros que variam de leve descamação com discreta perda de cabelo até grandes placas inflamatórias e pustulosas.^{2,3} A TC pode ser classificada clinicamente como tonsurante ou inflamatória. Enquanto a tonsurante se subdivide em microspórica e tricofítica, a inflamatória é subdividida em supurativa (ou *kerion*) e fávica²⁻⁴ (fig. 1).

O tratamento da TC baseia-se no uso de terbinafina ou griseofulvina, sem evidências clínicas para o uso de outros antifúngicos orais.^{1,4,5} A griseofulvina foi o primeiro fármaco eficaz usado no tratamento da TC e ainda é amplamente utilizada em locais com poucos recursos.^{1,5,6} Até agora, a terbinafina demonstrou bom perfil de segurança, e é considerada boa alternativa para o tratamento da *Tinea capitis* em crianças.⁴⁻⁷

Frente a essas possibilidades de tratamento da TC, realizou-se estudo observacional retrospectivo com a análise dos dados obtidos de prontuário de pacientes com

DOI do artigo original:

<https://doi.org/10.1016/j.abdp.2023.03.008>

☆ Como citar este artigo: Proença CGC, Carvalho GSM, Valinoto GCJ, Mayor SAS, Veasey JV. *Tinea capitis*: observations and clinical approach in a pediatric population of 99 cases. *An Bras Dermatol*. 2024;99:279-83.

☆☆ Trabalho realizado no Departamento de Dermatologia da Santa Casa de Misericórdia de São Paulo, Hospital Central, São Paulo, SP, Brasil.