

## Validando a Inteligência Artificial antes do Uso à Beira do Leito: ChatGPT-4 versus Mini em Conteúdo sobre Hipertensão Baseado em Diretrizes

*Validating Artificial Intelligence before the Bedside: ChatGPT-4 vs Mini in Guideline-Based Hypertension Content*

Cláudio Tinoco Mesquita,<sup>1,2,3,4</sup>  Fernanda Azevedo Silva,<sup>1,2</sup> Ana Beatriz Costa do Couto<sup>2,3</sup> 

Universidade Federal Fluminense,<sup>1</sup> Niterói, RJ – Brasil

Hospital Universitário Antônio Pedro, EBSERH, Health, Science & Education Lab,<sup>2</sup> Niterói, RJ – Brasil

Programa de Pós-Graduação em Ciências Cardiovasculares – Universidade Federal Fluminense,<sup>3</sup> Niterói, RJ – Brasil

Hospital Pró-Cardíaco,<sup>4</sup> Rio de Janeiro, RJ – Brasil

**Minieditorial referente ao artigo: Avaliação do ChatGPT-4.0 Versus ChatGPT-Mini na Geração de Conteúdo sobre Hipertensão Baseado em Diretrizes**

A integração da inteligência artificial (IA) na prática médica é uma realidade. Desde o suporte à tomada de decisão clínica e o desenho de estudos até ferramentas de chatbots e a produção de materiais educativos, os modelos de linguagem de grande escala (LLMs) são cada vez mais utilizados na prática clínica. Nesse contexto, estudos como o de Ataídes et al.,<sup>1</sup> que avaliam a performance de materiais educacionais baseados em LLMs, são fundamentais.

A hipertensão é uma das principais causas de morte prematura; sendo este um dos motivos que fazem a Organização Mundial da Saúde (OMS) estimular o desenvolvimento de campanhas de prevenção. Trata-se de uma doença que afeta múltiplos órgãos, causando grande impacto em todo o organismo. A condição é definida por pressão arterial sistólica  $\geq 140$  mmHg ou pressão arterial diastólica  $\geq 90$  mmHg, de acordo com as diretrizes da Sociedade Brasileira de Cardiologia.<sup>2</sup> O número de indivíduos diagnosticados com hipertensão aumenta a cada ano, e estudos indicam que há grande influência dos hábitos alimentares da população, que atualmente são caracterizados por alto consumo de alimentos ultraprocessados.<sup>3</sup>

O estudo que comparou ChatGPT-mini e ChatGPT-4.0 na geração de conteúdo educacional sobre hipertensão aborda um aspecto muito importante da IA na medicina: a qualidade e a segurança das informações direcionadas aos pacientes. Ao avaliar sistematicamente acurácia, completude, qualidade estrutural (EQIP), consistência e alinhamento com diretrizes clínicas, os autores realizaram uma avaliação abrangente dos conteúdos educacionais que foram gerados por IA.<sup>1</sup>

À primeira vista, comparar versões de modelos de LLMs pode parecer algo sensível ao tempo, em especial em um campo caracterizado por rápida evolução tecnológica. Na verdade, em fevereiro de 2026, os modelos de LLMs baseados no GPT-4 foram definitivamente descontinuados no ecossistema do ChatGPT, sendo substituídos por gerações mais recentes (GPT-5). No entanto, isso não diminui a relevância do estudo. Pelo contrário, reforça sua mensagem central.

Os autores devem ser parabenizados, pois o verdadeiro valor do trabalho não está em determinar qual modelo apresenta melhor desempenho, mas em demonstrar que nenhum conteúdo gerado por IA deve ser considerado confiável sem validação rigorosa.<sup>4</sup> Mesmo modelos avançados, capazes de produzir textos fluentes, coerentes e aparentemente confiáveis, podem gerar imprecisões, omitir informações críticas ou apresentar recomendações desalinhadas com diretrizes estabelecidas.<sup>5</sup> De forma importante, este estudo reforça um princípio fundamental que deve ser seguido independentemente do modelo ou versão de LLM: a responsabilidade pelo conteúdo oferecido aos pacientes é atribuição dos profissionais de saúde e as instituições — e não responsabilidade do algoritmo.<sup>4,6</sup> (Figura 1).

Por fim, nunca é excessivo reforçar, o uso ético da IA é essencial, especialmente quando LLMs são utilizados no cuidado ao paciente. Os pacientes devem ser informados quando o conteúdo for produzido com auxílio da IA, e as instituições devem estabelecer processos claros de revisão, aprovação e atualização periódica dos materiais educativos.

Em conclusão, embora as tecnologias continuem a evoluir, o princípio fundamental permanece inalterado: antes que a IA alcance o paciente, ela deve ser cuidadosamente testada, validada e contextualizada.

### Palavras-chave

Inteligência Artificial; ChatGPT; Hipertensão; Diretrizes Clínicas; Validação

**Correspondência:** Claudio Tinoco Mesquita •

Universidade Federal Fluminense, Hospital Universitário Antônio Pedro –

Health, Science & Education Lab – Av. Marquês do Paraná, 303.

CEP 24033-900. Niterói, RJ – Brasil

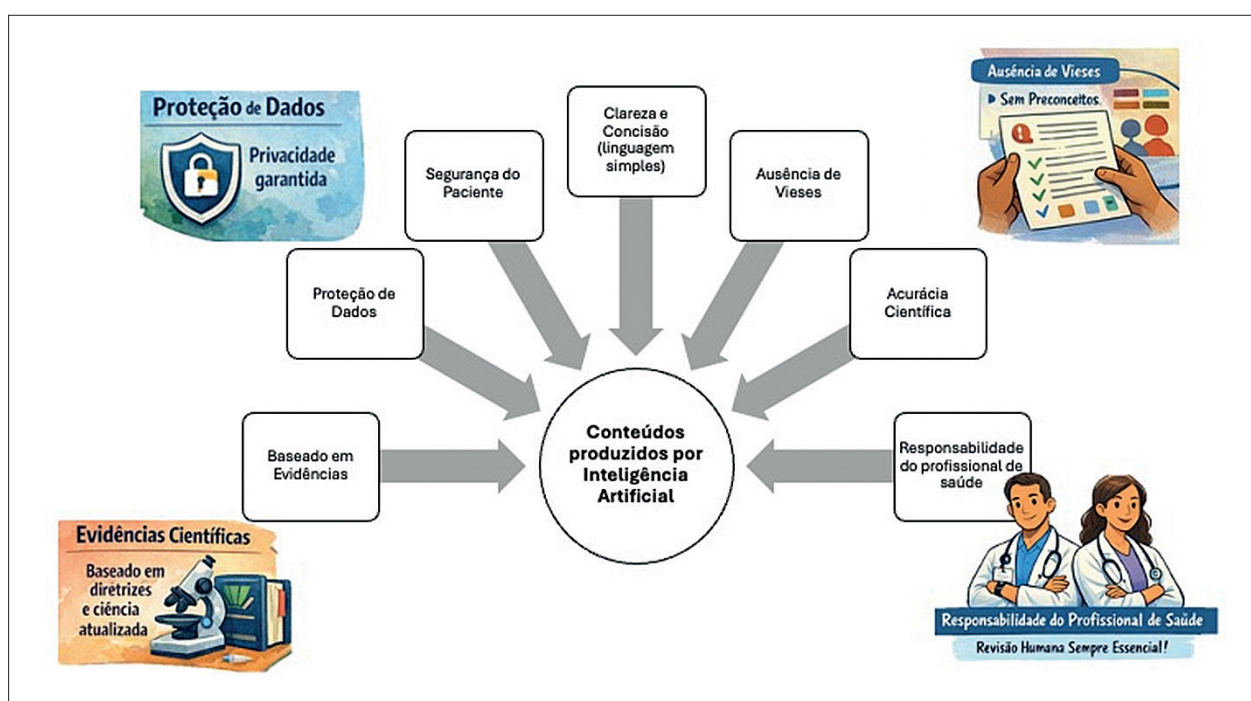
E-mail: claudiotinocomesquita@id.uff.br

Artigo recebido em 25/03/2026, revisado em 25/03/2026, aceito em 01/04/2026

**DOI:** <https://doi.org/10.36660/abc.20260254>

### Agradecimentos

Dr. Cláudio Tinoco recebe os seguintes fomentos: Bolsa de Produtividade do CNPq e Bolsa de Cientista do Estado do Rio de Janeiro pela FAPERJ.



**Figura 1** – Esta figura resume os principais domínios para avaliar conteúdos gerados por IA para pacientes: acurácia científica e base em evidências, revisão por especialista e responsabilidade, clareza e legibilidade, ausência de vieses, segurança do paciente e privacidade de dados. Em conjunto, esses elementos reforçam que os conteúdos produzidos por IA exigem validação sistemática e supervisão humana antes do uso clínico.

## Referências

1. Ataídes RJC, Campos MAG, Souza JVP, Rocha RC, Lacalle AA, Vieira CB, et al. Assessment of ChatGPT-4.0 versus ChatGPT-Mini in Generating Guideline-Based Hypertension Content. *Arq Bras Cardiol.* 2026;123(2):e20250498. doi: 10.36660/abc.20250498.
2. Brandão AA, Rodrigues CIS, Bortolotto LA, Armstrong ADC, Mulinari RA, Feitosa ADM, et al. Brazilian Guidelines of Hypertension - 2025. *Arq Bras Cardiol.* 2025;122(9):e20250624. doi: 10.36660/abc.20250624.
3. Alhuneafat L, Al Ta'ani O, Arriola-Montenegro J, Al-Ajloun YA, Naser A, Chaponan-Lavalle A, et al. The burden of Cardiovascular Disease in Latin America and the Caribbean, 1990-2019: An Analysis of the Global Burden of Disease Study. *Int J Cardiol.* 2025;428:133143. doi: 10.1016/j.ijcard.2025.133143.
4. Griffin AC, Javier SJ, Golding M, Runnels TW, Matthias MS, Shimada SL, et al. Designing Patient-Friendly Messages: Tutorial on Applying Human-Centered, Self-Determination Theory with AI Considerations. *J Med Internet Res.* 2025;27:e78173. doi: 10.2196/78173.
5. Wang C, Liu S, Yang H, Guo J, Wu Y, Liu J. Ethical Considerations of Using ChatGPT in Health Care. *J Med Internet Res.* 2023;25:e48009. doi: 10.2196/48009.
6. Mello MM, Char D, Xu SH. Ethical Obligations to Inform Patients about Use of AI Tools. *JAMA.* 2025;334(9):767-70. doi: 10.1001/jama.2025.11417.



Este é um artigo de acesso aberto distribuído sob os termos da licença de atribuição pelo Creative Commons