

Entre o bisturi e algoritmos: como a IA está redesenhando a Cirurgia

Between the scalpel and algorithms: how AI is redesigning Surgery

CRISTIANO XAVIER LIMA TCBC-MG¹ ; MARCOS ANDRÉ GONÇALVES² ; ARTHUR DE OLIVEIRA LIMA³ .

INTRODUÇÃO

A cirurgia sempre esteve na confluência entre ciência, tecnologia e habilidades humanas. Nas últimas décadas, o campo se expandiu muito além da sala de cirurgia: medicina perioperatória, tratamento multimodal do câncer, plataformas robóticas e transplantes complexos agora fazem parte da realidade cotidiana¹. Esse rápido progresso traz benefícios inegáveis, mas também gera um paradoxo: espera-se que os cirurgiões dominem um conjunto cada vez maior de conhecimentos, ao mesmo tempo em que ofereçam cuidados seguros e oportunos em situações mais desafiadoras². O resultado é uma lacuna cada vez maior entre o ritmo da inovação e a capacidade dos indivíduos e dos sistemas de saúde de absorvê-la. Ao mesmo tempo, a força de trabalho cirúrgica enfrenta demandas crescentes — desde treinar novas gerações em técnicas avançadas até lidar com as implicações éticas e econômicas das novas tecnologias³. Nesse cenário, o desafio não é mais o acesso à informação, mas a capacidade de transformar quantidades enormes de dados em decisões clinicamente significativas. A inteligência artificial, e particularmente os grandes modelos de linguagem (LLMs), surgiram como ferramentas disruptivas com o potencial de remodelar a forma como os cirurgiões aprendem, ensinam e praticam⁴.

O processamento de linguagem natural (NLP) tem sido explorado há muito tempo na medicina, desde a extração de informações estruturadas de registros eletrônicos de saúde até o apoio à documentação clínica

e à mineração de literatura^{5,6}. Essas aplicações anteriores, embora valiosas, limitavam-se a tarefas restritas e predefinidas. O recente surgimento de grandes LLMs representa um salto qualitativo neste campo, passando de sistemas baseados em regras para modelos versáteis, capazes de compreender o contexto, gerar narrativas coerentes e participar em diálogos interativos^{7,8}. Para o campo da cirurgia, essa evolução não é meramente incremental: os LLMs abrem a possibilidade de transformar a forma como os cirurgiões aprendem procedimentos complexos, ensinam as futuras gerações e integram evidências na prática diária⁹.

As raízes dos LLMs remontam aos avanços no aprendizado de máquina, particularmente no aprendizado profundo e nas arquiteturas transformadoras introduzidas em 2017, que permitiram o manuseio de longas sequências de texto com eficiência sem precedentes⁸. Ao contrário dos sistemas tradicionais de NLP estatísticos ou baseados em regras, os LLMs são pré-treinados em vastos conjuntos de textos heterogêneos e, em seguida, adaptados para tarefas específicas do domínio de aplicação, permitindo-lhes generalizar em diferentes contextos⁷. Os modelos de última geração — como o GPT-4¹⁰ e variantes ajustadas por instruções como o Flan-PaLM¹¹ — codificam não apenas estruturas linguísticas, mas também conhecimento factual e processual. Sua versatilidade reside na capacidade de processar dados não estruturados, gerar resumos, traduzir idiomas e simular padrões de raciocínio¹². No domínio clínico, esses recursos se traduzem em acesso mais ágil ao conhecimento, suporte educacional em tempo real

1 - Universidade Federal de Minas Gerais (Departamento de Cirurgia) - Belo Horizonte - MG - Brasil 2 - Universidade Federal de Minas Gerais (Departamento de Ciências da Computação) - Belo Horizonte - MG - Brasil 3 - Politecnico di Torino (Dipartimento di Automática e Informatica) - Turim - Itália

e comunicação aprimorada entre equipes multidisciplinares⁹. Para os cirurgiões, as vantagens incluem a capacidade de recuperar rapidamente as melhores práticas, redigir relatórios operatórios estruturados, apoiar a tomada de decisões com evidências sintetizadas e fornecer recursos de ensino adaptáveis que se alinham com o nível de especialização do aluno⁴.

A integração dos LLMs à medicina segue uma linha no tempo que reflete tanto a prontidão tecnológica quanto a adoção clínica. Os primeiros marcos na medicina geral incluem o uso de NLP para codificação automatizada na década de 1990⁵, os primeiros sistemas de apoio à decisão clínica na década de 2000¹³ e o sucesso de LLMs ajustados para instrução, como o Flan-PaLM11, na melhoria do desempenho em benchmarks médicos após sua introdução no início da década de 2020¹⁴. Desde 2022, artigos científicos demonstraram que modelos como o Flan-PaLM alcançam precisão de ponta em tarefas MultiMedQA, incluindo uma precisão superior a 67% em questões do tipo USMLE¹⁴. Na cirurgia, a adoção ficou um pouco atrás, mas está acelerando rapidamente: as explorações iniciais por volta de 2021 se concentraram na automação de notas operatórias¹⁵; em 2022, estudos de viabilidade demonstraram a utilidade dos LLMs na geração de listas de verificação perioperatórias¹⁶; e a partir de 2023, as primeiras plataformas educacionais começaram a integrar LLMs em módulos de simulação e treinamento cirúrgico¹⁷. Essa trajetória dupla ressalta uma mudança progressiva: do suporte administrativo na medicina geral para aplicações clinicamente relevantes e orientadas para procedimentos em cirurgia, com cada estágio exigindo validação rigorosa antes da integração na prática.

Os LLMs estão agora na vanguarda da IA médica e têm grande potencial no trabalho clínico, na educação e na pesquisa⁹. As barreiras à implementação imediata nesses três domínios representam oportunidades para um maior desenvolvimento que podem ser exploradas por desenvolvedores de LLMs e equipes de pesquisa independentes. Atualmente, o uso de LLMs ainda é limitado na medicina, por sua falta de precisão, atualidade, coerência e transparência, bem como por questões éticas^{18,4}. A tecnologia LLM pode, no entanto, ter um impacto substancial na forma como o trabalho médico é realizado, particularmente onde os riscos são

menores, onde os dados pessoais não são necessários e onde o conhecimento especializado não é necessário ou é fornecido pelo usuário¹⁹.

Nos últimos anos, algumas novas aplicações potenciais de grandes modelos de linguagem foram propostas para o contexto da cirurgia^{15,17}. Olhando para o futuro, uma das propostas mais promissoras é a integração de LLMs com modelos de visão-linguagem (VLMs). Embora os VLMs possam interpretar imagens e vídeos cirúrgicos, muitas vezes não têm a capacidade de contextualizar os seus resultados dentro dos fluxos de trabalho clínicos. Os LLMs podem aprimorar esses sistemas traduzindo padrões visuais complexos em narrativas clinicamente significativas, vinculando achados intraoperatórios a conhecimentos baseados em evidências e gerando feedback para o treinamento cirúrgico. Essa integração multimodal poderia transformar a tomada de decisões e a educação perioperatórias, mas também amplifica a necessidade de validação rigorosa para evitar vieses ou padronizações simplificadas demais.

Algumas propostas de VLMs foram recentemente apresentadas especificamente para o contexto cirúrgico, incluindo protótipos como SurgicalGPT²⁰, Surgical-VQA²¹, Surgical-VQLA²² e Surgical-LVLM²³. Estes sistemas combinam a extração de características visuais com LLMs para interpretar cenas cirúrgicas, classificar fases procedimentais ou responder a perguntas estruturadas sobre ferramentas e etapas. Embora essas abordagens demonstrem a viabilidade da IA multimodal em contextos cirúrgicos, elas permanecem em grande parte experimentais. A maioria dos modelos se concentra em tarefas de classificação restritas, em vez de produzir narrativas clinicamente úteis; o treinamento é restrito a alguns conjuntos de dados cirúrgicos e a transparência é limitada, com apenas um modelo de código aberto disponível para validação independente. Em contrapartida, os VLMs médicos, como Med-Gemini e GPT-4 Omni, são de código fechado e não divulgam claramente a extensão dos dados cirúrgicos em seu treinamento. Esse panorama ilustra tanto a promessa quanto a imaturidade dos VLMs na cirurgia: eles destacam a próxima fronteira da integração multimodal, mas também reforçam a necessidade de validação rigorosa, transparência e liderança cirúrgica antes que essas ferramentas possam entrar com segurança na prática clínica.

Ainda assim, a área carece de ensaios rigorosos e pragmáticos que validem essas ferramentas em ambientes cirúrgicos reais. Se os cirurgiões não assumirem a liderança na construção e validação de aplicações baseadas em LLMs, há um risco concreto de que essas tecnologias sejam desenvolvidas sem levar em conta as complexidades únicas do cuidado cirúrgico. Tais sistemas podem introduzir vieses, promover padronizações inseguras e ignorar variáveis intraoperatórias cruciais que não podem ser capturadas em conjuntos de dados baseados em texto.

Para além da sala de cirurgia, a ausência de liderança cirúrgica pode distorcer o uso dessas tecnologias na educação médica, seja simplifican-

do em excesso etapas procedimentais ou oferecendo orientações inadequadas a residentes e, ainda em última instância, comprometendo tanto a segurança do paciente quanto o desenvolvimento do profissional.

Como a cirurgia historicamente abraçou a inovação — da anestesia às técnicas minimamente invasivas e à robótica — a comunidade cirúrgica agora enfrenta uma nova responsabilidade: testar criticamente, aprimorar e orientar a integração dos LLMs na prática, na educação e na pesquisa. Se a cirurgia pretende permanecer na vanguarda da inovação médica, os cirurgiões não devem apenas adotar os LLMs, mas também liderar sua validação rigorosa e integração ética na prática clínica.

REFERÊNCIAS

1. Alderson D. The future of surgery. *Br J Surg*. 2019;106(1):9-10. doi:10.1002/bjs.11086.
2. McCulloch P. Innovation in surgery. *BMJ Surg Interv Health Technol*. 2019;1(1):e000021. doi:10.1136/bmjsit-2019-000021.
3. Shrimme MG, Dare A, Alkire BC, Meara JG. Catastrophic expenditure to pay for surgery worldwide: a modelling study. *Lancet Glob Health*. 2015;3 Suppl 2(0 2):S38-44. doi: 10.1016/S2214-109X(15)70085-9.
4. Patel S, Lam K. Large language models in surgery: applications and future directions. *Ann Surg Open*. 2023;4(2):e343. doi:10.1097/AS9.0000000000000343.
5. Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB. A general natural-language text processor for clinical radiology. *J Am Med Inform Assoc*. 1994;1(2):161-74. doi:10.1136/jamia.1994.95236141.
6. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? *J Biomed Inform*. 2009;42(5):760-72. doi:10.1016/j.jbi.2009.08.007.
7. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: *NAACL-HLT*. Minneapolis: ACL; 2019. p.4171-86.
8. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: *NeurIPS*. 2017. p.5998-6008.
9. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Tan GS, et al. Large language models in medicine. *Nat Med*. 2023;29:1930-40. doi:10.1038/s41591-023-02577-1.
10. OpenAI. GPT-4 technical report. *arXiv [Preprint]*. 2023. Available from: <https://arxiv.org/abs/2303.08774>
11. Chung HW, Hou L, Longpre S, Zoph B, Tay Y, Fedus W, et al. Scaling instruction-finetuned language models. *J Mach Learn Res*. 2024;25(70):1-53. doi: 10.48550/arXiv.2210.11416.
12. Thieme A, Nori A, Ghassemi M, Bommasani R, Andersen TO, Luger E. Foundation models in healthcare: opportunities, risks & strategies forward. In: *CHI Conference on Human Factors in Computing Systems*. ACM. 2023;1-15. doi:10.1145/3544549.3585897.
13. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence.

- JAMA. 2018;320(21):2199-200. doi:10.1001/jama.2018.17163.
14. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. *Nature*. 2023;620:172-80. doi:10.1038/s41586-023-06291-2.
 15. Dagli MM, Ghenbot Y, Ahmad HS, Chauhan D, Curlip RT, Wang P, et al. Development and validation of a novel AI framework using NLP with LLM integration for chart review. *Sci Rep*. 2024;14:26783. doi:10.1038/s41598-024-69965-4.
 16. Yu T, Li X, Zhou Y, Wu H, Sun S, Guo H, et al. Application of an artificial intelligence-based system for verification of perioperative safety. *Am J Transl Res*. 2024;16(7):3131-9. doi: 10.62347/PUUT2092.
 17. Satapathy S, Pai A, Kumar N, Rajesh A, Khan S, Pandey R, et al. Artificial intelligence in surgical education and training. *Int J Surg*. 2023;109:133-4. doi:10.1016/j.ijsu.2023.06.055.
 18. Jeblick K, Schachtner B, Dexl J, Mittermeier A, Stüber AT, Topalis J, et al. ChatGPT makes medicine easy to swallow. *Radiology*. 2022;307(2):e223312. doi:10.1148/radiol.223312.
 19. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How well does ChatGPT do when taking the medical licensing exams? *PLOS Digit Health*. 2023;2(2):e0000206. doi:10.1371/journal.pdig.0000206.
 20. Seenivasan L, Islam M, Kannan G, Ren H. SurgicalGPT. In: MICCAI 2023. Cham: Springer. 2023;281-90. doi:10.1007/978-3-031-43901-8_27.
 21. Seenivasan L, Islam M, Krishna AK, Ren H. Surgical-VQA. In: MICCAI 2022. Cham: Springer; 2022. p.33-43. doi:10.1007/978-3-031-16449-0_4.
 22. Bai L, Islam M, Seenivasan L, Ren H. Surgical-VQLA. In: ICRA 2023. IEEE; 2023. p.6859-65. doi:10.1109/ICRA48891.2023.10161234.
 23. Wang G, Li Y, Chen X, Zhang J, Zhao L, Xu H, et al. Surgical-LVLM. *arXiv [Preprint]*. 2024. Available from: <https://arxiv.org/abs/2405.1094>.

Disponibilidade de Dados

Os dados relacionados a este artigo estarão disponíveis mediante solicitação ao autor correspondente.

Recebido em: 25/09/2025

Aceito para publicação em: 06/11/2025

Conflito de interesses: não.

Fonte de financiamento: nenhuma.

Editor

Daniel Cacione

Endereço para correspondência:

Cristiano Xavier Lima

E-mail: cxlima@ufmg.br

