



Prompt engineering in medical education. Dissecting the new technological frontier in Digestive Surgery

VISUAL ABSTRACT

Prompt Engineering in Medical Education

The New Digital Stethoscope in the Era of Generative AI

THE CONTEXT



AI Capability

Passes USMLE
High Clinical Insight



User Skill

Generic Queries
Low Interaction Quality

"Prompt Engineering is the new digital stethoscope."

THE 4 LEVELS OF PROMPTING

4. Chain of Thought

3. Few-Shot (Examples)

2. Persona + Context

1. Direct Question

Anatomy of a Perfect Prompt

- Context: Scene
- Persona: Role
- Task: Objective
- Format: Output

RISKS & SAFEGUARDS



Hallucinations

Risk of phantom refs.
Must verify all citations.



Bias

Avoid male-default.
Explicitly request DEI parameters.



Human Judgment

The final, irreplaceable filter for patient safety.

CONCLUSION: AI processes data, but Clinical Judgment remains irreplaceable.

AUTHORS

Ozimo Pereira GAMA FILHO

Breno Dias LIMA RIBEIRO , Paulo KASSAB

CORRESPONDENCE

Ozimo Pereira Gama Filho.

Email: ozimo.gama@ufma.br

HOW TO CITE THIS ARTICLE

Gama Filho OP, Lima Ribeiro BD, Kassab P. Prompt engineering in medical education. Dissecting the new technological frontier in Digestive Surgery.

ABCD Arq Bras Cir Dig. 2026;39:e1936.

<https://doi.org/10.1590/0102-67202026000007e1936>

ARTICLE HIGHLIGHTS

- The anatomy of a "perfect prompt" rests on four fundamental pillars.
- The most clinically relevant impact of advanced prompting is the mitigation of catastrophic errors.
- Ultimately, Prompt Engineering is more than a technical skill; it is a new pedagogy.
- The success of medical education lies not merely in the transfer of facts, but in forging a mind capable of interrogating reality with precision.



Prompt engineering in medical education. Dissecting the new technological frontier in Digestive Surgery

Engenharia do Prompt na Educação médica. Dissecando a nova fronteira tecnológica na Cirurgia Digestiva

Ozimo Pereira GAMA FILHO¹ , Breno Dias LIMA RIBEIRO² , Paulo KASSAB³

ABSTRACT

Generative Artificial Intelligence (AI) has become a tangible reality in medical education, yet many struggle to interact effectively with these models. This editorial introduces Prompt Engineering as the modern “digital stethoscope”—a systematic approach to maximize AI potential. A structured prompt relies on four pillars (Context, Request, Persona, Format) and advances through four levels of complexity. Strategies like instructing the AI to “think step by step” mitigate logical errors in clinical tasks. However, users must remain vigilant against AI “hallucinations” and phantom citations by using verified databases. Ultimately, while AI processes vast data, human clinical judgment remains the irreplaceable filter for patient safety.

Headings: Artificial Intelligence. Medical Education. Engineering. Ethics, Medical. Computer-Assisted Instruction.

RESUMO

A Inteligência Artificial (IA) Generativa tornou-se uma realidade tangível na educação médica, contudo, muitos ainda têm dificuldade em interagir de forma eficaz com esses modelos. Este editorial apresenta a Engenharia de Prompt como o “estetoscópio digital” moderno — uma abordagem sistemática para maximizar o potencial da IA. Um prompt estruturado baseia-se em quatro pilares (Contexto, Requisição, Persona, Formato) e avança por quatro níveis de complexidade. Estratégias como instruir a IA a “pensar passo a passo” mitigam erros lógicos em tarefas clínicas. No entanto, os usuários devem permanecer vigilantes contra as “alucinações” da IA e citações fantasmas, utilizando bancos de dados verificados. Por fim, embora a IA processe vastas quantidades de dados, o julgamento clínico humano permanece como o filtro insubstituível para a segurança do paciente.

Descritores: Inteligência Artificial. Educação Médica. Engenharia. Ética Médica. Instrução por Computador.

INTRODUCTION

Generative Artificial Intelligence (AI) has transitioned from a futuristic promise to a tangible academic reality, evidenced by its ability to perform at or above the threshold of the United States Medical Licensing Examination (USMLE) without specialized training. However, a paradox persists in medical schools: while AI demonstrates surprising clinical insight, most students and faculty have yet to master the “language” required to interact with these models effectively. To understand the qualitative leap of this tool, one must observe its technical evolution: GPT-1 (2018) operated with 117 million parameters, whereas modern models process trillions, changing the paradigm from simple indexing to complex knowledge synthesis. For the surgeon, this means AI has evolved from a medical dictionary into a logic consultant, demanding instructions as precise as the cleavage plane in an oncological dissection.

In this context, Prompt Engineering emerges as the “digital stethoscope” of the modern era (Figure 1). As defined by Hes-ton et al., it is a systematic approach to communicating with AI to obtain high-precision results, maximizing the potential of Generative Language Models (GLMs)⁶.

The dialogue with AI is not a unidirectional query but a structured hierarchy of sophistication. The anatomy of a

“perfect prompt” (Figure 2) rests on four fundamental pillars, much like the technical rigor required in a cholecystectomy: Context (setting the scenario), General Request (the core objective), Persona/Role (defining the behavior, such as a “Professor of Surgical Gastroenterology”), and Output Format (e.g., comparative tables or structured FAQs). Based on frameworks adapted for medical education, prompts can be scaled in complexity from Level 1 (Direct Question), Level 2 (Persona and Context), Level 3 (Few-Shot/Example-based, e.g., providing a clinical template for mesenteric ischemia), to Level 4 (Component Decomposition or “Chain of Thought”)⁶. Recent systematic reviews highlight that well-structured prompts are essential for generating high-quality multiple-choice questions (MCQs) and clinical vignettes, ensuring the AI functions as a precision pedagogical tutor^{1,2,7}.

The most clinically relevant impact of advanced prompting is the mitigation of catastrophic errors. Instructing a GLM to “think step by step” utilizes its predictive nature to create an external “working memory,” significantly reducing errors in complex logical reasoning. In the surgical environment, this logic is what prevents iatrogenesis during the multi-level staging of gastric cancer or in complex pharmacological calculations^{5,7}. Furthermore, AI can be used to simulate realistic

¹Universidade Federal do Maranhão, Faculty of Medicine, Topographic Anatomy – São Luís (MA), Brazil.

²Universidade Federal do Maranhão, Faculty of Medicine, Department of Medicine – São Luís (MA), Brazil.

³Santa Casa of São Paulo, Medical School, Department of Surgery – São Paulo (SP), Brazil.

Correspondence: Ozimo Pereira Gama Filho. Email: ozimo.gama@ufma.br

Financial source: None.

Conflicts of interests: None.

Received: 02/01/2026. Accepted: 03/07/2026

Editor: Nelson Adami Andreollo

How to cite this article: Gama Filho OP, Lima Ribeiro BD, Kassab P. Prompt engineering in medical education. Dissecting the new technological frontier in Digestive Surgery. ABCD Arq Bras Cir Dig. 2026;39:e1936. <https://doi.org/10.1590/0102-67202026000007e1936>

“Virtual Patients”. For instance, generating a dynamic scenario to evaluate post-operative pain control strategies after bariatric surgery allows students to train clinical reasoning before real patient contact, echoing the trends observed in recent medical curricula⁵.

Despite its potential, the risk of “hallucinations” (Figure 3), where AI prioritizes stylistic fluency over factual accuracy, remains a critical concern for researchers. The generation of non-existent Vancouver-style citations (phantom citations) demands the integration of evidence-based tools (like Consensus or Elicit) and the strict verification of primary databases^{3,8}.

Furthermore, the integration of AI in medical education must address ethical imperatives. Poorly designed prompts can perpetuate biases. Therefore, instructions must explicitly include Diversity, Equity, and Inclusion (DEI) parameters, such as requesting cardiovascular or digestive cases in female patients to explore atypical presentations^{5,8}.

In the Brazilian public health context, where high-stakes exams like The National Residency Exam (ENARE) and The National Medical Training Assessment Exam (ENAMED) demand extreme temporal efficiency — such as creating personalized mnemonics for acute pancreatitis severity criteria

PROMPT ENGINEERING IN MEDICINE THE NEW STETHOSCOPE

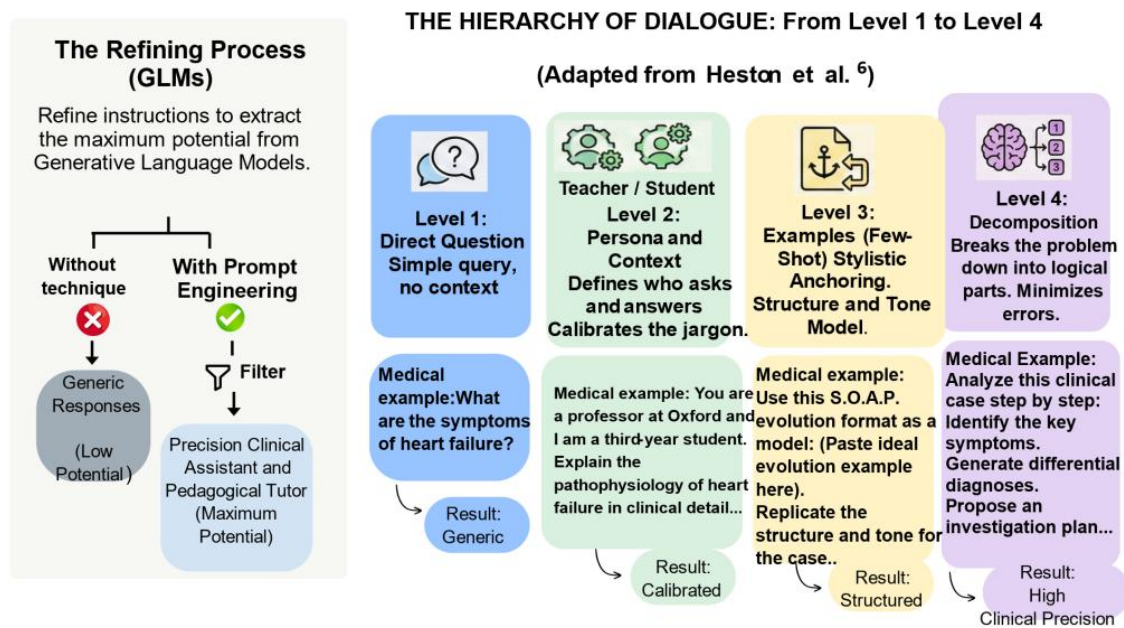


Figure 1. The perfect prompt.

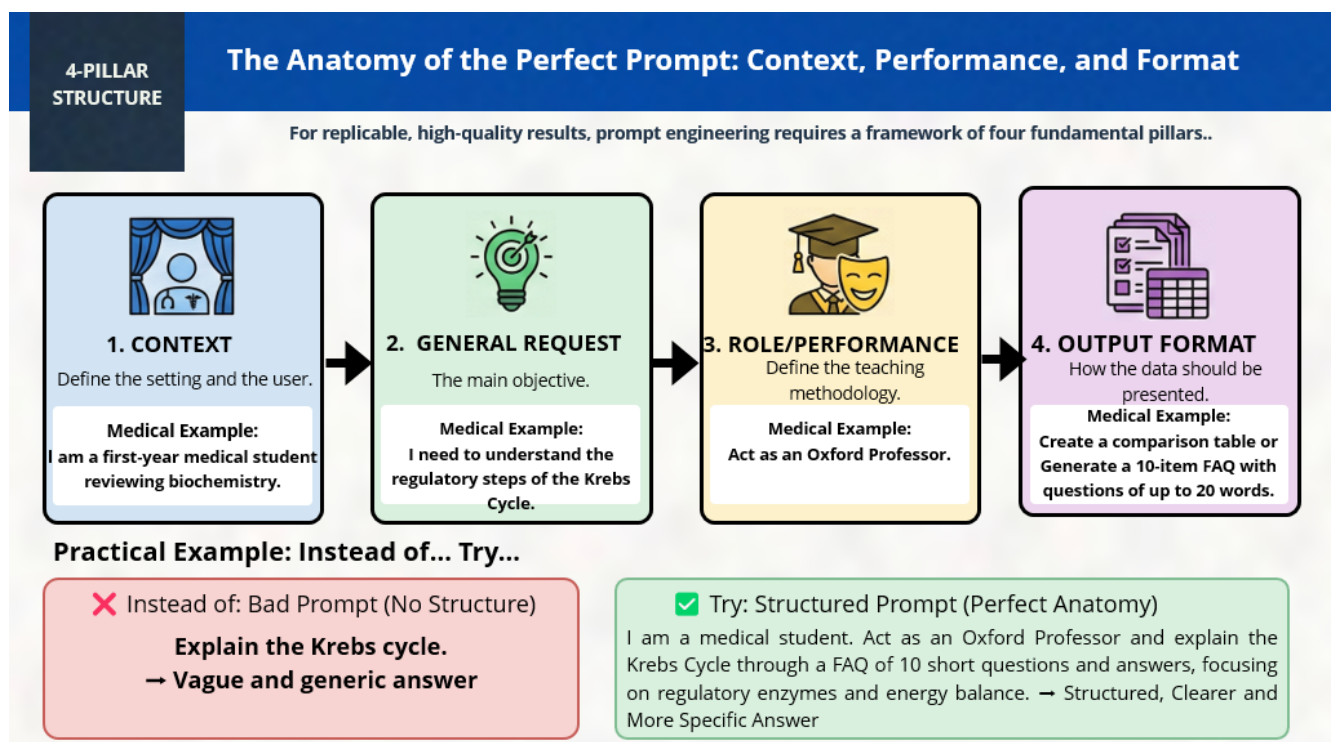


Figure 2. Prompt engineering.

BEWARE OF ACADEMIC "HALLUCINATIONS" AND PHANTOM CITATIONS

The tendency of AI to prioritize stylistic fluency over factual accuracy is a major risk for the researcher.

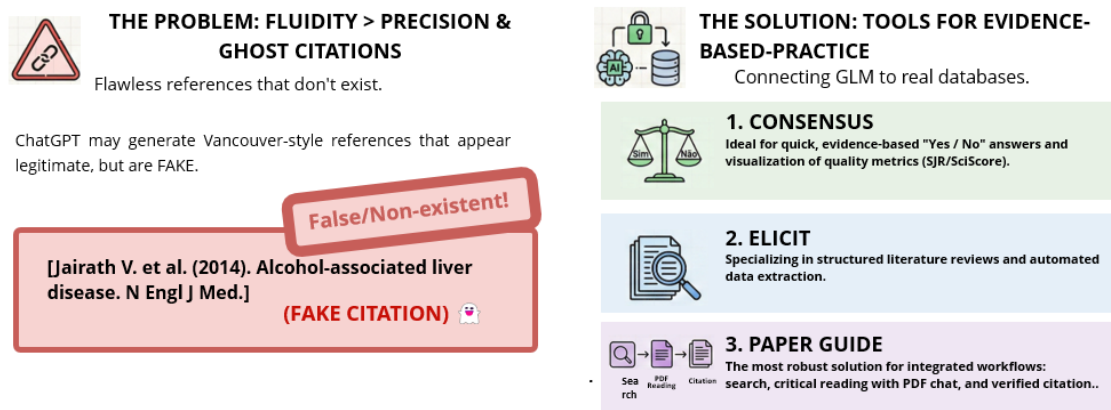


Figure 3. Hallucination prevention.

— the adoption of federated learning and AI-assisted reviews provides vital lessons for resource-constrained settings^{3,4,9}. This emphasizes that technological integration must be culturally and structurally aligned.

CONCLUSIONS

Prompt Engineering is more than a technical skill. It is a new pedagogy that forces us to be more critical thinkers. The quality of the AI's response is a direct reflection of the clarity of human thought. While AI can process vast amounts of data, human clinical judgment remains the final and irreplaceable filter. As adapted from the principles of William Osler, the success of medical education lies not merely in the transfer of facts, but in forging a mind capable of interrogating reality with precision. Those who fail to embrace AI will be at a disadvantage, but those who trust it blindly will be at risk.

AUTHORS CONTRIBUTION

OPGF: Conceptualization, Data analysis, Investigation, Methodology, Writing – original article. BDLR: Data analysis, Literature review, Methodology, Writing – original article. PK: Writing – original article.

DATA AVAILABILITY

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Al Shuraiqi S, AlZaabi A, Aal Abdulsalam A. Prompt engineering strategies for generating medical case-based MCQs with large language models: a multi-model comparative study. *Mach Learn Knowl Extr.* 2026;8(2):41. <https://doi.org/10.3390/make8020041>
2. Artsi Y, Sorin V, Konen E, Glicksberg BS, Nadkarni G, Klang E. Large language models for generating medical examinations: systematic review. *BMC Med Educ.* 2024;24(1):354. <https://doi.org/10.1186/s12909-024-05239-y>
3. Borges FT, Machado GM, Santana MA, Sancho KA, França GVA, Santos WP, et al. LLM-assisted scoping review of artificial intelligence in Brazilian public health: lessons from transfer and federated learning for resource-constrained settings. *Int J Environ Res Public Health.* 2026;23(1):81. <https://doi.org/10.3390/ijerph23010081>
4. Facanali Junior MR, Sousa Junior AHS, Marques CFS, Safatle-Ribeiro AV. Artificial intelligence-assisted colonoscopy for colorectal lesion detection: a case-control study on diagnostic accuracy and histopathological agreement. *ABCD Arq Bras Cir Dig.* 2025;38:e1898. <https://doi.org/10.1590/0102-67202025000029e1898>
5. Gordon M, Daniel M, Ajiboye A, Uraiby H, Xu NY, Bartlett R, et al. A scoping review of artificial intelligence in medical education: BEME Guide No. 84. *Med Teach.* 2024;46(4):446-70. <https://doi.org/10.1080/0142159X.2024.2314198>
6. Heston TF, Khun C. Prompt engineering in medical education. *Int Med Educ.* 2023;2(3):198-205. <https://doi.org/10.3390/ime2030019>
7. Riehm L, Nanji K, Lakhani M, Pankiv E, Hasanee D, Pfeifer W. The use of large language models in generating multiple choice questions for health professions education: a systematic review and network meta-analysis. *PLoS One.* 2026;21(1):e0340277. <https://doi.org/10.1371/journal.pone.0340277>
8. Masters K, MacNeil H, Benjamin J, Carver T, Nemethy K, Valanci-Aroesty S, et al. Artificial intelligence in health professions education assessment: AMEE Guide No. 178. *Med Teach.* 2025;47(9):1410-24. <https://doi.org/10.1080/0142159X.2024.2445037>
9. Tustumi F, Andreollo NA, Aguilar-Nascimento JE. Future of the language models in healthcare: the role of ChatGPT. *Arq Bras Cir Dig.* 2023;36:e1727. <https://doi.org/10.1590/0102-672020230002e1727>