Prompting for Healthcare Professionals: Enhancing Clinical Decision-Making with Artificial Intelligence

Citation: Alemanno A., Carmone M., Priore L. "Prompting for Healthcare Professionals: Enhancing Clinical Decision-Making with Artificial Intelligence" (2025) *Infermieristica Journal* 4(1): 33-39. DOI: 10.36253/if-3198

Received: December 16, 2024

Revised: February 2, 2025

Just accepted online: March 29, 2025

Published: March 31, 2025

Correspondence: Antonio Alemanno, Medical Physics, University Hospital "Policlinico Riuniti" of Foggia Email: alemannox@gmail.com

Data Availability Statement: All relevant data are within the paper and its Supporting Information files. This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination, and proofreading process, which may lead to differences between this version and the Version of Record.

Competing Interests: The author(s) declare(s) no conflict of interest.

Antonio Alemanno[1], Michele Carmone[1], Leonardo Priore[2]

[1]*Health Physics - Azienda Ospedaliero-Universitaria Policlinico Foggia, Italy*
[2]*Department of Emergency Medicine, Emergency Room - Policlinico University Hospital of Foggia, Italy*

Abstract

Introduction. Generative Artificial Intelligence (AI), specifically through Large Language Models (LLMs), is progressively reshaping clinical documentation, decision support, patient education, and research synthesis in healthcare. Despite significant benefits, these models pose challenges such as inaccuracies (hallucinations) and inherent biases. This paper highlights prompt engineering as an emerging and critical skill for healthcare professionals and demonstrates how structured prompting techniques can improve the reliability, clinical relevance, and ethical compliance of AI-driven applications.

Methods. A systematic review of recent literature was conducted to present structured prompt engineering methodologies specifically tailored to healthcare settings. Advanced prompting techniques, including chain-of-thought reasoning, zero-shot and few-shot prompting, and self-consistency strategies, were examined.

Results. The proposed structured approach encompasses clear objective definition, precise contextualization, integration of domain-specific knowledge, iterative refinement, and ethical risk mitigation. Practical guidelines are provided for designing prompts suitable for clinical scenarios, such as diagnostic decisions, patient-specific therapeutic protocols, and administrative tasks. Notably, advanced techniques such as chain-of-thought reasoning and self-consistency effectively reduce inaccuracies and enhance clinical decision-making.

Discussion and Conclusion. The structured integration of prompt engineering optimizes clinical decision-making and supports adherence to evidence-based practices. Incorporating prompt engineering into healthcare educational programs and fostering interdisciplinary collaboration are crucial for the responsible implementation of generative AI. These advancements have far-reaching implications for improving clinical effectiveness, enhancing patient care quality, and elevating the standards of healthcare education and professional practice.

Keyword: Artificial Intelligence, Prompt Engineering, Healthcare Professionals

Introduction

Generative AI models, such as Large Language Models (LLMs), are demonstrating revolutionary potential in healthcare. These advanced tools enable significant improvements in clinical documentation, patient education, decision support, and the synthesis of scientific research[1,2].

However, challenges associated with the use of LLMs include issues such as "hallucinations," where outputs may appear plausible but are inaccurate. To address these issues, prompt engineering emerges as an essential skill for healthcare professionals. By strategically designing and iteratively refining the inputs provided to AI, it is possible to ensure that the models produce outputs that are relevant and aligned with clinical guidelines[3,4].

Prompt Engineering in Healthcare: a structured approach

Clear definition of objectives and outcomes

The first step is to precisely define the purpose of the prompt, which can range from summarizing clinical data to formulating diagnoses or extracting specific patient information[5,6]. It is also essential to clearly specify the desired output format, such as a text, table, or procedural guide. For instance, a prompt focused on therapeutic management might request a comparative table of pharmacological options, while another might generate a structured list of differential diagnoses or simple tables for scheduling activities[2,6].

Example of a prompt with a defined output:

Create an Excel table with a detailed weekly schedule of radiological exams, including columns for the day, time slot, type of exam, patient ID, patient age, and clinical priority ("Urgent," "Routine," "Post-operative follow-up"). Ensure technical timing is respected (e.g., 30 minutes for exams with contrast, 15 minutes for those without), and assign urgent priorities to the earliest available time slots each day. Distribute routine and follow-up exams evenly throughout the week while avoiding scheduling conflicts and overloading specific days. Include a color-coded system for visual clarity: red for urgent, yellow for routine, and green for follow-up exams.

Prompt contextualization

Integrating contextual information, such as demographic data and the patient's medical history, enhances the accuracy of AI responses. For example, a prompt evaluating therapeutic options for a diabetic patient should include details like age, comorbidities, and current treatments[8,9]. Furthermore, aligning the prompt with clinical guidelines and evidence-based practices enhances the precision of diagnostic coding and patient stratification activities[1,4].

Example of a prompt with a guideline-based output:

You are a dietitian tasked with creating a weekly meal plan for a 50-year-old man with type 2 diabetes. Consider the following:

- Guidelines: Follow the 2024 American Diabetes Association (ADA) nutritional recommendations, which include a balanced diet with 45-60% of daily calories from carbohydrates, healthy fats, and lean proteins.

- Nutritional needs: Distribute 1,800 kcal per day across three main meals and two snacks. Prioritize low-glycemic-index carbohydrates, such as whole grains and legumes, while avoiding refined sugars.

- Food preferences: The patient prefers simple, quick-to-prepare meals. Avoid red meat, but include fish and tofu.

- Goal: Manage blood glucose levels, improve body weight, and enhance insulin sensitivity.

- Required format: Provide the plan in a daily table format, specifying breakfast, lunch, dinner, and snacks. Include precise portions and a weekly shopping list.

- Additional instructions: Add practical recommendations for meal preparation and suggest food substitutions (e.g., oatmeal instead of packaged cereals).

Iterative refinement of prompts

Prompt engineering is a dynamic and iterative process that requires constant testing and refinement. Through feedback from clinical and technical experts, ambiguities, redundancies, or biases in prompts can be identified and resolved, improving their effectiveness and precision. Prompts can range from simple structures, such as the zero-shot approach—where the model is not provided with specific examples [9,10]—to more advanced strategies. The zero-shot method is ideal for obtaining quick answers to direct questions and proves particularly effective in training healthcare professionals[11]. However, gradually adding details, instructions, and constraints makes prompts more specific and contextualized.

Using examples, known as few-shot prompting, further clarifies the expected task, facilitating the model's learning and improving its accuracy. Advanced techniques such as chain-of-thought prompting or self-consistency prompting can be employed to manage complex tasks, progressively increasing sophistication and ensuring greater control over the quality and reliability of the output [3,6].

Since prompts must be designed to minimize the risk of hallucinations and ensure adherence to ethical principles, such as patient privacy protection and fairness, techniques like consistency checks and iterative prompts are essential for verifying output accuracy and reducing the margin of error [4,8].

Example of an iterative refinement prompt:

Initial Prompt

- Provide instructions for the management of pressure ulcers.

First Refinement

- Develop a detailed protocol for the management of pressure ulcers, including:
  - Techniques for assessment and staging of pressure ulcers
  - Procedures for cleaning and debridement
  - Selection and application of appropriate dressings
  - Strategies to reduce pressure and prevent further injuries
  - Frequency of dressing changes and monitoring of healing progress

Second Refinement

- Create a customized protocol for the management of pressure ulcers for a 70-year-old patient with limited mobility and diabetes mellitus. Specify:
  - Initial assessment of the ulcer, including staging according to the guidelines of the National Pressure Injury Advisory Panel (NPIAP)
  - Cleaning and debridement plan, considering potential complications related to diabetes
  - Selection of advanced wound care products (e.g., hydrocolloids, foams, alginates)
  - Interventions to relieve pressure, such as the use of cushions or pressure-relieving mattresses
  - Education for the patient and caregivers on the prevention of new ulcers and the importance of glycemic control
  - Regular monitoring of healing progress and adjustment of the care plan as needed based on the patient's condition

Example of zero-shot prompt

The technique involves asking a language model to per-form a specific task without providing any explanatory examples in the prompt, relying solely on the model's general understanding:

A 70-year-old patient with a history of hypertension and type 2 diabetes pre-sents to the emergency department complaining of chest pain and shortness of breath. Describe the priority nursing actions you would take in this situa-tion.

Example of few-shot prompt

In this technique, a few explanatory examples (usually one to five) are provided within the prompt to guide the model in performing a specific task, without requiring additional training:

Analyze the results in the provided file and assign an urgency level (High, Medium, Low) for each blood parameter.

For example:

1. *A patient with hemoglobin at 6 g/dL is classified as High urgency.*
2. *A patient with hemoglobin at 12 g/dL is classified as Low urgency.*

Example of Chain-of-thought prompt

It encourages the model to explicitly articulate, step by step, the logical reasoning required to solve a problem or complete a task. This strategy helps improve the con-sistency and accuracy of responses, especially for tasks requiring complex or multi-step decision-making process-es:

Develop a nursing intervention plan for a 75-year-old patient with dyspnea and fluid retention, explaining step by step the clinical reasoning behind each recommended action.

Example Self-consistency prompt:

This prompt is an advanced prompting technique for lan-guage models that relies on generating multiple respons-es to the same input, followed by identifying the most consistent or frequent response among those generated. This strategy enhances the reliability of the result by se-lecting the most representative or logical option, mitigat-ing uncertainty or inconsistencies in individual outputs:

You are an expert nurse specializing in multiple and independent clinical evaluations. For each clinical scenario I present to you, you will need to gen-erate several reasoned responses and subsequently analyze them to identify the most coherent and reliable solution.

Evaluation Procedure:

1. Generate 3-5 independent approaches to the clinical situation.

2. Compare the responses by identifying:

- Convergences among the proposed solutions.

- Significant discrepancies.

- The most solid clinical rationale.

Evaluation Criteria:

- Patient safety.

- Clinical evidence.

- Professional guidelines.

Scenario Example:

An elderly patient with pneumonia, persistent fever, and oxygen saturation fluctuating between 89-92%.

Below is a concise checklist (Table 1.) summarizing the key steps for effective prompt engineering in healthcare:

Table 1. Checklist of key steps for effective prompt engineering in healthcare.

Step	Key Actions
1. Define Clear Objectives	- Identify the purpose of the prompt (e.g., diagnosis, treatment planning, documentation).
	- Specify the desired output format (text, table, structured list).
2. Contextualize the Prompt	- Include relevant information such as medical history, demographics, and patient preferences.
	- Align the prompt with evidence-based guidelines.
3. Refine Iteratively	- Test the initial prompt and gather feedback from experts.
	- Remove ambiguities and enhance details and instructions.
	- Integrate advanced techniques (e.g., chain-of-thought, few-shot prompting).
4. Mitigate Ethical Risks	- Ensure privacy protection and fairness in responses.
	- Use consistency checks to minimize errors and biases.
5. Evaluate and Monitor	- Compare the generated output with clinical criteria and guidelines.
	- Conduct accuracy checks and revise the prompt as necessary.

Regarding point 4 of the checklist, in the development of AI systems in healthcare, data protection principles must be applied from the design phase of any prompts that access databases[12].

Contextual prompt design for long inputs

The ability of LLMs to process long inputs significantly impacts context-based prompt design. For instance, shorter contextualization prompts can help users effectively establish the necessary context before deploying a main prompt. Additionally, directing file inputs exclusively to LLMs capable of handling extended text ensures optimal performance and accuracy. These strategies have been shown to enhance output quality and alignment with user objectives[13].

Example of contextualization with short prompts for nursing documentation:

Imagine a nurse needs to use an LLM to generate a detailed care plan based on the nursing records of a patient with a chronic wound. Instead of inputting the entire nursing file in one step, the nurse could use shorter contextualization prompts as follows:

- Contextualization Prompt 1:

"Summarize the nursing notes on the patient's wound care over the past week, including wound size, type of dressing used, and any signs of infection."

- Contextualization Prompt 2:

"Based on the summarized wound care notes, identify

potential complications and nursing interventions already in place."

- Main Prompt:

"Using the provided information, create a detailed care plan for the next week, including dressing changes, monitoring for infection, pain management, and patient education."

This strategy ensures the LLM processes the nursing records incrementally, maintaining focus on key aspects like wound care and intervention strategies. It also allows the nurse to verify that critical details from the nursing notes are accurately captured before requesting a comprehensive care plan.

Strategies for evaluating variability in LLM outputs

Large Language Models (LLMs) like ChatGPT can produce substantially different responses even when presented with the same prompt. This variability arises from factors such as the probabilistic nature of language generation and differences in the model's training process. For instance, studies have reported accuracy fluctuations of up to 10% in deterministically configured LLMs when identical inputs are processed repeatedly, highlighting the challenge of ensuring consistent outputs[14]. Additionally, prompt sensitivity has been identified as a key determinant of model performance, further underscoring the need for robust and systematic evaluation frameworks[15].

In healthcare, these variations necessitate structured comparison methods to ensure that the selected output is both reliable and clinically valid. Standardized reporting and evaluation metrics have been introduced to address variability in LLM applications[16]. Furthermore, consistent human evaluation methodologies tailored to healthcare scenarios are essential for improving the practical utility of AI-generated solutions[17].

By adopting such frameworks, healthcare professionals can mitigate risks, make better-informed decisions, and ensure the clinical feasibility of LLM-generated outputs. When comparing outputs from different models to identify the most reliable results, recommended strategies include standardizing prompts, generating multiple outputs per model for comparison, involving multidisciplinary teams to review clinical relevance and accuracy, and validating results against evidence-based guidelines and established decision-support systems.

Techniques for mitigating user-induced biases and enhancing data quality

To address user-induced biases, prompting techniques such as self-critique prompts can be employed. For instance, asking the model to critically evaluate its previous response[18] encourages it to identify inconsistencies and refine the output through a checklist of detected errors. Similarly, using divergent reasoning chains, such as the Divergent Chain of Thought method[19], can enhance accuracy by requiring the model to generate and compare alternative reasoning pathways before reaching a conclusion.

For intrinsic biases, strategies like negative prompts and mirror-consistency have proven particularly effective. Negative prompts explicitly instruct the model to exclude specific assumptions (e.g., gender stereotypes), while mirror-consistency[20] leverages discrepancies in "minority" outputs to identify potential uncertainties and recalibrate confidence in final responses.

Additionally, ensuring high data quality is essential for bias mitigation, as inconsistencies or inaccuracies in training data can reinforce biases in AI-generated outputs.

To further enhance data quality and reduce reliance on potentially biased or inaccurate information from the open web, AI systems can be integrated into "controlled data ecosystems", where prompting is applied exclusively to curated sources, such as medical guidelines, peer-reviewed literature, or institutional databases. Platforms like *NotebookLM*[21] or *ChatGPT's custom knowledge retrieval*[22] offer structured environments where AI can process and generate responses based on pre-selected, authoritative documents, minimizing the risks associated with unverified data and improving the reliability of AI-driven decision support in healthcare.

Integrating these techniques into practice helps healthcare professionals mitigate biases and ensure that LLM-generated information aligns with ethical standards, particularly in terms of data quality, as highlighted in the WHO guidelines on AI for health[23].

Conclusions

The applications of prompt engineering span various areas, including clinical decision support, such as generating differential diagnoses, formulating therapeutic plans, and analyzing complex clinical data to enhance diagnostic accuracy and support evidence-based decisions[2,24]; patient education, by producing personalized, clear, and accessible materials useful for managing chronic diseases[4,7]; and administrative tasks, where AI systems can automate appointment scheduling, documentation, and data management, reducing administrative burdens and freeing up time for clinical activities [4,6].

Prompt engineering thus represents an emerging and crucial skill to fully leverage the potential of generative AI. By designing clear, contextualized, and ethically sound prompts, healthcare professionals can guide AI models to produce reliable and relevant outcomes, reducing the risks of errors and biases[1,4]. This skill transforms AI into a tool for improving clinical decision-making, streamlining workflows, and fostering more effective communication between patients and professionals.

In the future, integrating prompt engineering into training programs and fostering interdisciplinary collaborations will be essential to further refine these techniques and ensure their responsible implementation[2,8], especially if we manage to bring together experts from different healthcare professions on collaborative platforms for AI innovation in healthcare.

References

1. Shah, K., Xu, A. Y., Sharma, Y., Daher, M., McDonald, C., Diebo, B. G., & Daniels, A. H. (2024). Large language model prompting techniques for advancement in clinical medicine. Journal of Clinical Medicine, 13(17), 5101. https://doi.org/10.3390/jcm13175101
2. Patil, R., Heston, T. F., & Bhuse, V. (2024). Prompt engineering in healthcare. Electronics, 13, 2961. https://doi.org/10.3390/electronics13152961
3. Meskó, B. (2023). Prompt engineering as an important emerging skill for medical professionals: Tutorial. Journal of Medical Internet Research, 25, e50638. https://doi.org/10.2196/50638
4. Nazi, Z. A., & Peng, W. (2024). Large language models in healthcare and medical domain: A review. Informatics, 11, 57. https://doi.org/10.3390/informatics11030057
5. Habib, M. M., Hoodbhoy, Z., & Siddiqui, M. A. R. (2024). Knowledge, attitudes, and perceptions of healthcare students and professionals on the use of artificial intelligence in healthcare in Pakistan. PLOS Digital Health, 3(5), e0000443. https://doi.org/10.1371/journal.pdig.0000443
6. Hendawi, S., Kanan, T., Elbes, M., Mughaid, A., & AlZu'bi, S. (2024). Automated prompt engineering pipelines: Fine-tuning LLMs for enhanced response accuracy [Preprint]. SSRN. https://doi.org/10.2139/ssrn.5004054
7. Chen, C. J., Liao, C. T., Tung, Y. C., & Liu, C. F. (2024). Enhancing healthcare efficiency: Integrating ChatGPT in nursing documentation. Studies in Health Technology and Informatics, 316, 851–852. https://doi.org/10.3233/SHTI240545
8. Abhari, S., Fatahi, S., Saragadam, A., Chumachenko, D., & Pelegrini Morita, P. (2024). A road map of prompt engineering for ChatGPT in healthcare: A perspective study. Studies in Health Technology and Informatics, 316, 998–1002. https://doi.org/10.3233/SHTI240578
9. Cai, C. J., Wang, H., Wang, M., Agrawal, R. K., & Bero, N. (2024). Prompting is all you need: LLMs for systematic review screening. MedRxiv. https://doi.org/10.1101/2024.06.01.24308323
10. Zaghir, J., Naguib, M., Bjelogrlic, M., Névéol, A., Tannier, X., & Lovis, C. (2024). Prompt engineering paradigms for medical applications: Scoping review. Journal of Medical Internet Research, 26, e60501. https://doi.org/10.2196/60501
11. Huang, M., & Sangi-Haghpeykar, H. (2023). Prompt engineering in medical education. International Medical Education, 2(3), 198–205. https://doi.org/10.3390/ime2030019
12. Garante per la protezione dei dati personali. Decalogo per la realizzazione di servizi sanitari nazionali attraverso sistemi di Intelligenza Artificiale. 2023 Oct 10
13. Hsieh CJ, Si S, Yu F, Dhillon IS. Automatic engineering of long prompts. Findings of the Association for Computational Linguistics: ACL 2024. Bangkok, Thailand: Association for Computational Linguistics; 2024. p. 10672–85.
14. Atil B, Chittams A, Fu L, Ture F, Xu L, Baldwin B. LLM Stability: A detailed analysis with some surprises. arXiv preprint arXiv:2408.04667. 2024.
15. Zhuo J, Zhang S, Fang X, Duan H, Lin D, Chen K. ProSA: Assessing and understanding the prompt sensitivity of LLMs. Findings of the Association for Computational Linguistics: EMNLP 2024; 1950–76. Miami, Florida, USA: Association for Computational Linguistics; 2024.
16. Ho CN, Tian T, Ayers AT, et al. Qualitative metrics from the biomedical literature for evaluating large language models in clinical decision-making: a narrative review. BMC Med Inform Decis Mak. 2024;24:357. https://doi:10.1186/s12911-024-02757-z
17. Tam TYC, Sivarajkumar S, Kapoor S, Stolyar AV, Polanska K, McCarthy KR, et al. A literature review and framework for human evaluation of generative large language models in healthcare. arXiv preprint arXiv:2405.02559. 2024.
18. Zhang W, Shen Y, Wu L, Peng Q, Wang J, Zhuang Y, et al. Self-Contrast: Better Reflection Through Inconsistent Solving Perspectives. arXiv preprint arXiv:2401.02009. 2024.
19. Puerto H, Chubakov T, Zhu X, Tayyar Madabushi H, Gurevych I. Fine-Tuning with Divergent Chains of Thought Boosts Reasoning Through Self-Correction in Language Models. OpenReview. 2024 Sep 24 [modified 2024 Dec 5]. https://openreview.net/forum?id=u4whlT6xKO
20. Huang S, Ma Z, Du J, Meng C, Wang W, Lin Z. Mirror-Consistency: Harnessing Inconsistency in Majority Voting. arXiv preprint arXiv:2410.10857. 2024.
21. Google Research. NotebookLM: AI-powered research and writing tool [Internet]. Google; 2023 [cited 2025 Feb 7]. Available from: https://notebooklm.google
22. OpenAI. ChatGPT: AI language model with custom knowledge retrieval [Internet]. OpenAI; 2023 [cited 2025 Feb 7]. Available from: https://openai.com/chatgpt
23. World Health Organization. Regulatory considerations on artificial intelligence for health. Geneva: WHO; 2023. p. 26-31. Available from: https://iris.who.int/handle/10665/373421
24. Neha, F., Bhati, D., Shukla, D. K., & Amiruzzaman, M. (2024). ChatGPT: Transforming Healthcare with AI. AI, 5(4), 2618-2650. https://doi.org/10.3390/ai5040126