Cogitare Enfermagem

ORIGINAL ARTICLE

Responses and accuracy of artificial intelligence systems to prompts on pressure injury management

HIGHLIGHTS

1. The accuracy of the AIs was compared via prompts on pressure injury.
2. The accuracy of the AIs was 31.9% with 23 correct answers.
3. Hybrid prompting showed higher accuracy than zero-shot.
4. Gemini® showed the highest accuracy for pressure injury management.

Letícia Pereira Pires[1]
Maithê de Carvalho e Lemos Goulart[1]
Fernanda Garcia Bezerra Góes[1]
Yonara Cristiane Ribeiro[1]
Luíza Lissonger Costa Guimarães[1]
Lediane Farias de Sousa[1]

ABSTRACT

Objective: To compare the responses and accuracy of different Artificial Intelligences based on a set of prompts for the management of pressure injury. **Method:** A descriptive, comparative study of the responses of Artificial Intelligences, carried out between October and November 2025, in two stages: 1) Study of prompts and 2) Test to obtain the responses. The following were evaluated: Perplexity AI, Grok AI, Blackbox AI, OpenAI ChatGPT-4.0®, Gemini®, and Claude AI, randomly divided into the hybrid prompting and zero-shot groups. The responses were organized in an Excel spreadsheet and analyzed descriptively. **Results:** The overall accuracy of the artificial intelligences was 31.9%. The hybrid prompting group reached 36.1% and the zero-shot obtained 27.7%. The Gemini® AI demonstrated the highest accuracy and ChatGPT-4.0® the worst. **Conclusion:** The accuracy presented indicates that the safest recommendation is the hybrid prompt model, integrating artificial intelligence with human clinical judgment.

DESCRIPTORS: Generative Artificial Intelligence; Large Language Models; Nursing, Team; Clinical Reasoning; Pressure Ulcer.

HOW TO REFERENCE THIS ARTICLE:
Pires LP, Goulart MCL, Góes FGB, Ribeiro YC, Guimarães LLC, de Sousa LF. Responses and accuracy of artificial intelligence systems to prompts on pressure injury management. Cogitare Enferm [Internet]. 2026 [cited "insert year, month and day"];31:e102398en. Available from: https://doi.org/10.1590/ce.v31i0.102398en

[1]Universidade Federal Fluminense, Departamento de Enfermagem, Rio das Ostras, RJ, Brasil.

INTRODUCTION

Pressure injuries (PI) are a frequent and global complication, with a prevalence that can reach 72.5%, varying according to the context[1]. Since 2013, the Patients for Patient Safety program has included the reduction of PI as its sixth goal, reinforcing the importance of measures that improve care and promote patient safety[2].

The National Pressure Injury Advisory Panel (NPIAP) defines PI as localized damage to the skin or underlying tissue, usually over bony prominences or areas subjected to pressure, shear, or chemical, physical, and biological factors[3]. They are classified into stages ranging from one (non-blanchable erythema) to four (full-thickness tissue loss with exposure of deep structures), in addition to unstageable injury, deep tissue injury, medical device-related injury, and mucosal membrane injury[4].

The occurrence of PI in institutionalized patients has been associated with care failures, such as inadequate repositioning, absence of physical examination, shortage of human and material resources, as well as insufficient mastery of dressings, wound care, and new technologies[2]. Scientific updating, combined with the nurse's technical skill, is essential for the appropriate choice of dressing, contributing to the quality of care and the prevention of complications[3].

Health technology has expanded in recent decades, with emphasis on Artificial Intelligence (AI), capable of rapidly processing large volumes of data and supporting personalized decisions[5]. The growing integration of AI into health services indicates a trend toward expanded use, making it essential to recognize its potential to enhance knowledge and practices. In this scenario of accelerated transformation, discussing the advantages and limitations of these models strengthens a critical stance toward the use of AI in health care[6].

In health care, AI has been introduced mainly as a tool to support clinical decision-making in complex situations, such as the identification of risk factors, prognosis definition, and the diagnostic accuracy of PI[7]. A German study reported the use of AI by nurses for the recognition, classification, reduction, prediction, and risk stratification of PI, highlighting its high degree of specificity[8], showing that AI not only complements routine clinical practices but also demonstrates its applicability in complex health scenarios.

Two main AI models are described: predictive AI, aimed at predicting outcomes from algorithms fed by large volumes of data; and generative AI, which articulates algorithms capable of creating content, including text, images, audio, among others[9]. Large Language Models (LLM) are AI systems trained with large volumes of data to recognize patterns and generate responses in natural language. When optimized through structured approaches, such as integration with knowledge bases and clinical guidelines, LLMs can show superior performance and alignment with guidelines and recommendations, including in the nursing context[10-11]. Chatbots (interaction spaces) powered by generative AI are consolidating themselves as a support resource, offering personalized responses to users[6].

AI can develop personalized recommendations and approaches considering the patient's therapeutic scenario; however, these are responses given from commands called prompts and sophisticated algorithms. It also makes it possible to identify clinical patterns by analyzing extensive datasets[9], providing real-time information. Nevertheless, it should be noted that even the most refined models may not capture the full complexity of professionals' clinical expertise in the management of pressure injuries or chronic wounds[10-11].

The integration of this technology into nursing practice requires a cautious and conscious approach by the professional, and it is essential that its use reflects its potential benefits, limitations, and risks to patient safety[6]. However, the use of AI

should be understood as a support tool, and not as a substitute for clinical judgment and professional decision-making, and should be used to enhance care[5,10-11].

In this sense, the literature points to the need to train AI with a large volume of data, such as the LLMs; however, in care practice the professional uses the most accessible tool, without inserting structured data. Thus, there is a knowledge gap that needs to be filled in order to understand how these AIs behave when this is not done beforehand. The use of AI in everyday practice, therefore, must be combined with methodological rigor, the use of reliable data, and the requested prompt, ensuring its reliability and use as a facilitating resource in the management of PI, assisting in the choice of the most appropriate dressing and contributing to the integrity of the patient's skin.

Therefore, the objective of this study is to compare the responses and accuracy of different Artificial Intelligences from a set of prompts for the management of pressure injury.

METHOD

A descriptive and comparative study of the responses of Artificial Intelligences based on a set of prompts, carried out between October and November 2025, in two stages: 1) Study of prompts and 2) Test to obtain the AI responses.

The following AIs were used: Perplexity AI, Grok AI, Blackbox AI, OpenAI ChatGPT-4.0®, Gemini®, and Claude AI. The AIs were selected because they offer platforms whose terms of use allow the submission and analysis of PI images, in addition to free indexing. The AIs Microsoft Copilot, DeepSeek, and Meta AI were excluded because they do not read images according to their terms of use.

The AIs were randomly divided into two groups: the hybrid prompting model, articulating different techniques (few-shot, role-prompting, chain-of-thought, and chain-of-verification), and the zero-shot model of prompt engineering[12]. According to the Pan American Health Organization, the term prompt refers to the instructions provided to an AI system, with a brief question or detailed request, which specifies the tone, context, format, and desired objective, whose clarity and contextual grounding directly influence the quality of the generated response[13].

The hybrid prompting and zero-shot techniques were chosen, seeking to address, in a detailed and contextualized manner, the AI response to the nurse's decision-making needs for the management of PI and, based on their potential, to achieve precision and reliability in the responses[14].

The hybrid prompting technique integrates different prompt techniques, combining strengths to improve performance with more precise responses aligned with the objectives, which requires more complex instructions[15]. The combination occurs through few-shot, which provides examples to guide response patterns[14]; role prompting, which assigns a specific role to the model; chain-of-thought, which leads it to make its step-by-step reasoning explicit[14]; and chain-of-verification, structured in four stages: baseline response, verification plan, checks, and final response, increasing rigor and reliability[16]. The zero-shot strategy, in turn, consists of obtaining a response from a single initial command, without offering prior examples or additional information[14].

For data collection, which took place over two days in November 2025, each model was fed a previously selected image, corresponding to an unstageable PI according to the NPIAP consensus (Figure 1), inserted directly into the AIs using accounts (email and login) created exclusively for the study, in order to minimize biases arising from previous prompts in the individual user profile. The generated responses were recorded in the project profile and saved as a digital file through screenshots.

Figure 1. Image of unstageable pressure injury selected by the study team. Rio das Ostras, RJ, Brazil, 2025

Source: Feridas Crônicas - Recurso Educacional sobre Prevenção e Manejo da Lesão por Pressão (2020)[17].

The unstageable PI is characterized by full-thickness skin loss, whose actual depth and extent cannot be measured due to coverage by necrotic or devitalized tissue. Only after its removal does complete assessment become possible, generally revealing a stage 3 or 4 PI[18-19].

The randomization by drawing lots occurred before feeding the AI with the image, in a process in which a member of the research team wrote the names of the AIs on pieces of paper and placed them in a box. The drawing began with the hybrid prompting group and, after the first, the groups were alternated until the drawing was completed.

The group that received the zero-shot prompt was composed of Perplexity AI, Grok AI, and Blackbox AI, which received the following instruction: "Answer the following instrument based on the image." The aforementioned instrument was developed for this study and consists of 12 items written in the format of statements (Figure 2). The types of dressing listed in the checklist were defined based on the most frequent options used by nurses, according to the literature[18-20].

1. I classify this lesion as:
( ) Stage 1 ( ) Medical device-related pressure injury
( ) Stage 2 ( ) Deep tissue pressure injury
( ) Stage 3 ( ) Mucosal membrane pressure injury
( ) Stage 4 ( ) Unstageable

2. I identify that the image shows the following affected structures:
( ) Epidermis ( ) Cartilage
( ) Dermis ( ) Ligament
( ) Hypodermis ( ) Bone
( ) Fascia ( ) Tendons
( ) Muscle ( ) No structure affected
( ) Joint

3. Do I identify the following tissues in the lesion?
( ) Granulation tissue ( ) Fibrin
( ) Slough ( ) Necrotic tissue
( ) Eschar ( ) I identify none
( ) Epibole

4. I identify exudate in the lesion:
( ) Yes ( ) Maybe
( ) No

5. The wound edge shows signs of:
( ) Maceration ( ) Hyperemia
( ) Epibole ( ) No change identified

6. The periwound skin shows:
( ) Preserved integrity ( ) Erythema
( ) Edema ( ) Desquamation
( ) No change identified

7. I would cleanse this PI with:
( ) Distilled water ( ) Ringer's solution ( ) I would not cleanse it with any of these materials
( ) Chlorhexidine ( ) PHMB
( ) Soap ( ) Iodine
( ) 0.9% saline solution ( ) Hydrogen peroxide

8. I choose the following item(s) as the primary dressing for this PI dressing:
( ) Essential fatty acids ( ) Foam dressings ( ) Activated charcoal
( ) Sugar ( ) Barrier cream/spray ( ) Calcium alginate
( ) Hydrofiber ( ) Silver-impregnated dressings ( ) Negative pressure
( ) Hydropolymer ( ) Collagenase ( ) Papain
( ) Hydrogel ( ) Hydrocolloid

9. The main objective of the dressing I chose is:
( ) Maintain wound moisture ( ) Remove devitalized tissue
( ) Control bacterial load ( ) Promote granulation
( ) Control pain

10. About debridement:
( ) Would not perform it ( ) Would perform mechanical debridement
( ) Would perform enzymatic debridement ( ) Would perform surgical debridement

11. I choose the following item(s) as the secondary dressing for this PI dressing:
( ) Dry gauze ( ) Hydrocolloid ( ) Bandage
( ) Cotton ( ) Hydrofiber ( ) Micropore tape
( ) Film dressing ( ) Hydrocolloid ( ) Adhesive tape

12. I changed the dressing at the following frequency:
( ) Daily ( ) Every two days
( ) Every three days ( ) Only when saturated

Figure 2. Instrument developed for the AI group with zero-shot prompt. Rio das Ostras, RJ, Brazil, 2025

Source: The authors (2025).

The group composed of the AIs OpenAI ChatGPT-4.0®, Gemini®, and Claude AI received the hybrid prompting technique, with the prompt created for this study: *"You are a nurse specialized in wound dressings with experience in the assessment, classification, and management of pressure injuries. Your target audience is a patient presenting an integumentary lesion. Your objective, based on the attached image, is to classify the pressure injury, identify affected structures and tissues present, identify whether exudate is present, indicate how to perform the cleansing of this lesion, indicate the primary and secondary dressing, and recommend the frequency of dressing changes. The response format must be structured text, with an informative, professional, and concise tone. If the image is insufficient for a safe decision, explicitly state the limitations and describe which additional clinical data are necessary (e.g., vital signs, comorbidities, signs of infection, and exudate). Always use "suggestion"-type language when there is uncertainty and avoid stating a definitive medical diagnosis. To reach the conclusion, present the reasoning step by step and list the checks performed (at least 3), for example: correspondence with classification criteria, verification of signs of infection, compatibility of the dressing with the exudate, and verification of risks of maceration or trauma."*

The expected responses, considering both prompts, were defined based on the updated literature[18-19,21-22] on the subject and are presented in Chart 1.

Chart 1. Expected AI responses on the management of PI. Rio das Ostras, RJ, Brazil, 2025

(continue)

Item	Response	Justification
1. I classify this injury as:	Unstageable	Corresponds to an unstageable PI, with full-thickness skin loss and tissue damage whose depth cannot be determined because it is covered by slough or eschar. This covering prevents staging. After removal of the devitalized tissue, the injury may be reclassified as stage 3 or 4 PI, according to the depth and extent of the damage[16].
2. I identify that the image presents the following affected structures:	Dermis and epidermis	Epidermis and dermis as structures affected by full-thickness loss, involving both the superficial layer and the innermost layer of the skin[17].
3. I identify the following tissues in the injury	Necrotic tissue, slough, and granulation	Presence of necrotic tissue (eschar), slough, and granulation tissue, evidencing the coexistence of devitalized areas and areas in the process of repair[17].
4. I identify exudate in the injury:	No	No presence of exudate is evidenced, since there are no visible signs on the wound bed[17].
5. The edge of the injury presents signs of:	Epibole	Presents epibole at the injury edge, characterized by the rolled edge of the wound[17].
6. The skin around the injury presents:	Erythema	Area of erythema, characterized by redness of the skin resulting from increased local blood flow in the areas around the injury[17].
7. I would perform the cleansing of this PI with:	Running water or saline solution (0.9% saline) and Polyhexamethylene Biguanide (PHMB)	0.9% saline solution or PHMB, which are non-cytotoxic solutions, preserve fibroblasts and granulation tissue, and favor healing[17]. Running water is also a safe and low-cost alternative, especially in home care[19].

Chart 1. Expected AI responses on the management of PI. Rio das Ostras, RJ, Brazil, 2025

(conclusion)

Item	Response	Justification
8. I choose as the primary dressing for this PI dressing the following item(s):	Papain, hydrogel, no dressing, or gauze moistened with 0.9% saline.	Papain 10% favors chemical debridement. Papain at 2–4% aids tissue repair[20]. Hydrogel, to maintain a moist environment. Gauze moistened with 0.9% saline, in a low-resource context. Adhesive film, enzymes, or, in specific cases, no dressing, for necrosis. The film softens the eschar, while the enzymes break it down. For dry, intact eschar, if debridement is not part of the therapeutic plan, not applying a dressing is acceptable[17].
9. The main objective of the dressing I chose is:	To remove devitalized tissue, promote granulation, and maintain moisture	Removal of non-viable tissue, to maintain adequate moisture, prevent and control infections and odors, promote granulation, cleansing, and pain relief[17].
10. Regarding debridement:	Surgical, autolytic, mechanical, or enzymatic debridement	Debridement to remove non-viable tissue, reduce the risk of infection, and favor healing. Performed in a surgical, autolytic, mechanical, or enzymatic manner[17].
11. I choose as the secondary dressing for this PI dressing the following item(s):	Dry gauze and microporous adhesive tape	Dry gauze provides adequate protection and covering for the primary dressing, while the microporous adhesive tape performs the function of fixing the dressing to the skin[17].
12. I performed the dressing change at the following frequency:	Daily	To monitor the progression of healing, identify changes in the wound bed, and adjust the therapeutic approach[17].

Source: The authors (2025).

The responses collected from each AI were entered into a Microsoft Excel® spreadsheet and analyzed descriptively, being dichotomized as "correct" and "incorrect." The following were calculated: 1) the percentage of correct answers for each item of the instrument; 2) the total percentage of correct answers; and 3) the percentage of correct answers by AI group. It should be noted that items 1, 4, and 12 have only one correct alternative. In each question, only the alternatives that fully corresponded to the expected statements were considered "Correct"; all others were classified as "Incorrect." Responses presented with the "Maybe" option were also categorized as "Incorrect."

For the universe of correct answers, the 12 items of the instrument and the six AIs analyzed were considered, totaling 72 possible correct answers. For the overall accuracy of the AIs, the total number of correct answers was summed and divided by the universe of possible correct answers. For the calculation of the overall accuracy of the items, the correct answers for the item were summed and divided by the universe of possible correct answers for the item, considering the number of AIs, which was six. In the calculation of accuracy by group, the total number of correct answers in each group was summed and divided by the universe of possible correct answers, which was 36 in each. To calculate which AI was the most accurate, the total number of correct answers presented was divided by the total number of items of the instrument.

Submission of the study to the Research Ethics Committee was waived, since the research did not involve human beings, in accordance with Law No. 14,874/2024.

RESULTS

The overall accuracy of the AIs analyzed was 31.9% with 23 correct answers. However, the percentage of the hybrid prompting group was 36.1% with 13 correct answers and 23 errors in total, demonstrating slightly higher accuracy than the zero-shot group, which was 27.7% with 10 correct answers and 26 errors in total (Table 1).

Table 1. Percentage of correct answers according to the instrument for the six AI models. Rio das Ostras, RJ, Brazil, 2025

Item	Correct answers	Errors	Item accuracy	Zero-Shot Correct Answers	Hybrid Prompting Correct Answers
Item 1	2	4	**33.3%**	0	2
Item 2	0	6	**0.0%**	0	0
Item 3	1	5	**16.6%**	0	1
Item 4	1	5	**16.6%**	1	0
Item 5	0	6	**0.0%**	0	0
Item 6	3	3	**50.0%**	2	1
Item 7	6	0	**100.0%**	3	3
Item 8	1	5	**16.6%**	0	1
Item 9	2	4	**33.3%**	0	2
Item 10	6	0	**100.0%**	3	3
Item 11	0	6	**0.0%**	0	0
Item 12	1	5	**16.6%**	1	0
Overall accuracy (%): 31.9%				27.7%	36.1%

Source: The authors (2025).

Among the items analyzed, only two showed 100% correct answers: item 7, referring to the cleansing of the injury, and item 10, related to debridement, indicating consensus among all AIs on these aspects. In contrast, items fundamental to the assessment of the injury recorded no correct answers, such as item 2 (affected structures), item 5 (injury edge), and item 11 (secondary dressing), evidencing the models' inability to identify elements essential to the management of PI. In the remaining items, variability was observed between the two types of prompts, with specific differences in performance.

Among the AIs analyzed, Gemini® showed the highest accuracy with six (50.0%) correct answers, followed by Claude AI with five (41.6%) correct answers, Perplexity AI with four (33,3%) correct answers and Grok AI and Blackbox AI with three (25.0%) each. The worst accuracy was presented by ChatGPT-4.0® with two (16.6%) correct answers. The results demonstrated heterogeneity in the responses.

The hybrid prompting group showed higher accuracy in the identification and management of PI (Figure 3). Regarding the identification of PI, the zero-shot group had no correct answers concerning classification, affected structures, and injury edge.

Blackbox AI provided a correct response regarding the absence of exudate, and Perplexity AI and Grok AI in the recognition of erythema on the surrounding skin. The hybrid prompting group, in turn, was not accurate in the responses regarding affected structures, exudate, and injury edge. Both Gemini® and Claude AI correctly answered the classification of PI, with Gemini® also locating the erythema on the surrounding skin and Claude AI correctly describing the tissues present in the image.

Regarding the management of PI, cleansing with 0.9% saline and the need for debridement were unanimously indicated among the six AIs. The zero-shot group was not accurate in the items referring to the primary dressing, the objective of the dressing, and the secondary dressing. In the dressing-change item, only Perplexity AI answered that the change should occur daily. In the hybrid prompting group, the items of secondary dressing and dressing change were not answered correctly by any of the three AIs. Gemini® stood out by also getting the items of primary dressing and objective correct, and Claude AI in the objective.

LESION IDENTIFICATION							
Prompt	**AI**	**Classification**	**Affected structures**	**Tissues**	**Exudate**	**Wound edge**	**Periwound skin**
Zero-Shot	*Perplexity AI*	Stage 4	Epidermis, dermis, hypodermis, fascia, and muscle	Granulation tissue and necrotic tissue	Yes	Hyperemia	Erythema
	Grok AI	Stage 4	Dermis, hypodermis, fascia, muscle, bone	Slough, fibrin, and necrotic tissue	Yes	Hyperemia	Erythema
	BLACKBOX AI	Stage 4	Epidermis, dermis, hypodermis, and fascia	Slough and necrotic tissue	No	Hyperemia	Preserved integrity
Hybrid Prompting	*OpenAI Chat GPT-4.0®*	Stage 3 or 4	Epidermis, dermis, and subcutaneous tissue	Necrotic tissue and granulation tissue	Low or absent	Erythema	Hyperemia
	Gemini®	Unstageable	Epidermis, dermis, and subcutaneous tissue	Necrotic tissue and granulation tissue	Could not be reliably assessed	Hyperemia	Erythema
	Claude AI	Unstageable	Epidermis, dermis, and (probable subcutaneous tissue)	Necrotic tissue, granulation tissue, and slough	Could not be adequately assessed	Hyperemia	Hyperemia
MANAGEMENT OF THE PRESSURE INJURY							
Prompt	**AI**	**Cleansing**	**Primary dressing**	**Objective**	**Debridement**	**Secondary dressing**	**Dressing change**
Zero-Shot	*Perplexity AI*	0.9% saline solution	Hydrofiber, foam dressings, silver-impregnated dressing, and calcium alginate	Remove devitalized tissue, control bacterial load, and promote granulation	Surgical or enzymatic debridement	Film dressing and hydrocolloid	Daily
	Grok AI	0.9% saline solution	Hydrofiber, silver-impregnated dressing, collagenase, calcium alginate	Control bacterial load	Surgical debridement	Bandage and micropore tape	Every two days
	BLACKBOX AI	0.9% saline solution	Hydrofiber, hydropolymer, foam dressings, collagenase, hydrogel, and hydrocolloid	Remove devitalized tissue and promote granulation	Enzymatic debridement	Dry gauze, film dressing, bandage, micropore tape, and adhesive tape	Every three days
Hybrid Prompting	*OpenAI Chat GPT-4.0®*	0.9% saline solution	Hydrogel, silver dressing, povidone-iodine, silver alginate; keep eschar intact if dry and stable	Did not state an objective	Surgical debridement	Absorbent foam	48–72 hours; if infected, change daily
	Gemini®	0.9% saline solution	Hydrogel	Remove necrotic tissue	Autolytic debridement	Silicone-bordered foam dressing	24–48 hours
	Claude AI	0.9% saline solution	Hydrogel with or without calcium alginate	Maintain a moist environment and protect the wound	Surgical or enzymatic debridement	Sterile gauze secured with micropore tape, transparent film, polyurethane foam	24–48 hours

Figure 3. AI responses on the identification and management of PI. Rio das Ostras, RJ, Brazil, 2025

Legend: red - error in the response; green - correct answer.

Source: The authors (2025).

DISCUSSION

The findings of this study show that, although the AIs evaluated recognized fundamental and consensual practices in the management of PI, such as cleansing with saline solution and the indication of debridement, the overall accuracy can be considered low, since it did not even reach half of the possible correct answers. Despite the satisfactory performance of these tools in less complex tasks, important limitations were observed, including inconsistencies in interpretations and the possibility of hallucinations occurring, which compromises the reliability of the generated responses[7].

Hybrid prompting showed slightly higher accuracy when compared to zero-shot, including greater caution in the responses. In the item referring to exudate, for example,

two models recognized the limitations of the image and presented their responses in a suggestive tone when there were insufficient elements for definitive statements. The literature reinforces that the analysis should not be restricted to a single type of prompt, since elaborate and structured prompts tend to favor refined responses and more complete analyses[23].

Prompt engineering can increase accuracy in AI systems applied to health, as demonstrated by a European study that observed an increase from 80.1% to 99.6% when comparing queries performed with simple commands and more elaborate commands; however, this gain was accompanied by longer response time and an impact on user satisfaction[24]. It is reinforced that, in the clinical context, it is also necessary to validate other criteria, such as system speed, transparency with the AI, and usability of the prompting technique[25]. Thus, validation and evaluation studies become essential to verify whether the AI presents adequate performance for evidence synthesis and contributes safely to professional practice[26].

There is a gap in the literature, especially regarding the scarcity of references that address, in a thorough and systematic manner, prompting techniques applied to the management of PI. This gap restricts comparisons with the literature and evidences a field still under development. However, other clinical fields have been addressed internationally, such as a study developed in Sweden to evaluate the performance of LLMs in identifying medication errors, demonstrating that internal checking mechanisms (critic model) increase the reliability of the AI response[27], and a Korean study that compared different prompt engineering strategies, identifying the Chain-of-Thought method as producing more consistent responses in the therapeutic context of psychiatric nursing[28].

Regarding the indications of primary dressings, the AIs presented distinct possibilities. Although calcium alginate, foams, and silver dressings are compatible with wounds involving full-thickness loss, these options did not correspond to the injury analyzed[20]. Although the reference instrument did not include hydrofiber as a primary dressing, the AI models frequently cited it. The international literature recognizes this option as adequate because it favors autolytic debridement, indicated for removing devitalized tissue[29], compatible with the small amount of slough observed in the image.

AIs have the potential to support health professionals, increasing clinical efficiency and patient safety[1]. However, human competencies such as clinical reasoning, decision-making, and interpretation remain irreplaceable. Thus, AI recommendations must be evaluated by the professional before being applied[7]. Although it is pertinent to problematize the possible causes of the high number of errors of certain AIs, such as ChatGPT-4.0®, it is not possible, based on the data of this study, to accurately infer the determining factors of these failures. Therefore, the use of AI in the management of PI should be understood as a complementary tool, preserving the nurse's autonomy in decisions.

It is worth highlighting that there is a pressing need to incorporate artificial intelligence literacy into the training and continuing education of health professionals, as highlighted by a systematic review that investigated the level of AI literacy among health professionals and students, in which half of the included studies found very low literacy among both[30]. In this sense, it is not enough merely to use AI, but to develop competencies to formulate clear, contextualized, and clinically relevant questions, in addition to interpreting and validating the generated results. The absence of this mastery can lead to imprecise, mistaken, or potentially unsafe responses, even when apparently reliable models are used. Thus, AI literacy emerges as an essential component for the safe, ethical, and effective use of these technologies in the management of pressure injuries and in other care contexts.

This study has limitations regarding the single performance of each query to the AI, which limits the assessment of the variability of the responses, since generative models can produce different results in each interaction. Furthermore, the analysis was conducted on a restricted number of AI models and specific versions available during the research period, which may not reflect the performance of paid tools or other platforms.

This study did not evaluate inter-rater agreement, the consistency of the models, or the response time of each AI. Thus, the accuracy rate of each AI may be influenced by different interpretations by specialists and health professionals regarding what would be considered a correct response, which introduces variability in the evaluation and presents itself as another limitation of the study. Furthermore, the consistency of the models' responses was not measured, considering that an AI may seem reliable and produce stable responses and, even so, present systematic errors from a clinical perspective, which represents a risk in care contexts. The study also did not consider the response time of the models, a relevant aspect for clinical applicability, since faster models, even if less precise, may be more feasible and present better usability in the workflow.

Thus, these are points that direct future studies to broaden the notion of safety beyond accuracy, also incorporating dimensions such as consistency, usability, and efficiency in the real context of care. It is therefore recommended that, in the evaluation of the application of AI models in health, joint analyses of accuracy and consistency be carried out, especially in the multimodal context of the management of pressure injuries, in which interpretive complexity is high.

CONCLUSION

The accuracy of the AIs related to the management of PI, of 31.9%, was considered low; however, Gemini® demonstrated the highest accuracy and ChatGPT-4.0® the worst. Although the artificial intelligences were not originally developed for the health field, the accuracy presented indicates that the safest recommendation is the hybrid prompting model, integrating artificial intelligence with human clinical judgment.

By evidencing both potentialities and relevant limitations, it becomes clear that AI should be understood as a complementary tool in nursing practice, requiring responsibility and conscious use to ensure that its application strengthens care, without compromising patient safety and preserving the central role of the health professional in decision-making regarding the management of injuries.

In a scenario marked by the growing use of these models, it is essential to broaden the debate on the responsible integration of AI into care and to signal the need for future investigations to advance in the development of more robust methodologies. In this sense, it becomes essential to explore prompting strategies in different care contexts, contributing to the scientific and ethical improvement of the use of these technologies in nursing. Such an approach can prevent the indiscriminate adoption of tools whose applicability in health still requires further investigation and validation.

REFERENCES

1. Pott FS, Meier MJ, Stocco JGD, Petz FFC, Roehrs H, Ziegelmann PK. Pressure injury prevention measures: overview of systematic reviews. Rev Esc Enferm USP [Internet]. 2023 [cited 2025 Aug 15];57:e20230039. Available from: https://doi.org/10.1590/1980-220X-REEUSP-2023-0039en

2. Fontes TLA, de Oliveira BGRB, De Oliveira MF, da Silva MA, da Silva ARG, Pires BMFB, et al.

Development of a checklist to prevent pressure injuries in patients with COVID-19. Rev enferm UFPE on line [Internet]. 2024 [cited 2025 Aug 15];18(1):e257602. Available from: https://doi.org/10.5205/1981-8963.2024.257602

3. Macêdo SM, Bastos LLAG, Oliveira RGC, Lima MCV, Gomes FCF. Selection criteria for primary dressings in the treatment of pressure ulcers in hospitalized patients. Cogitare Enferm [Internet]. 2021 [cited 2025 Aug 25];26:e74400. Available from: http://dx.doi.org/10.5380/ce.v26i0.74400

4. de Araújo CAF, Pereira SRM, de Paula VG, de Oliveira JA, de Andrade KBS, de Oliveira NVD. Evaluation of the knowledge of nursing professionals in the prevention of pressure ulcer in intensive care. Esc Anna Nery [Internet]. 2022 [cited 2025 Aug 25];26:e20210200. Available from: https://doi.org/10.1590/2177-9465-EAN-2021-0200

5. Kotp MH, Ismail HA, Basyouny HAA, Aly MA, Hendy A, Nashwan AJ, et al. Empowering nurse leaders: readiness for AI integration and the perceived benefits of predictive analytics. BMC Nursing [Internet]. 2025 [cited 2025 Aug 25];24(56):1-13. Available from: https://doi.org/10.1186/s12912-024-02653-x

6. Dal Sasso GTM, Lanzoni GMM, Alvarez AG, Barra DCC, Barbosa SFF. Potential contribution of Chatgpt® to learning about septic shock in intensive care. Texto Contexto Enferm [Internet]. 2024 [cited 2025 Aug 25];33:e20230184. Available from: https://doi.org/10.1590/1980-265X-TCE-2023-0184en

7. Vitorino LM, Yoshinari Junior GH, Lopes-Júnior LC. Artificial intelligence in nursing: advancing clinical judgment and decision-making. Rev Bras Enferm [Internet]. 2025 [cited 2025 Nov 19];78(4):e780401. Available from: https://doi.org/10.1590/0034-7167.2025780401

8. Seibert K, Domhoff D, Bruch D, Schulte-Althoff M, Fürstenau D, Biessmann F, et al. Application scenarios for artificial intelligence in nursing care: rapid review. J Med Internet Res [Internet]. 2021 [cited 2025 Aug 22];23(11):e26522. Available from: https://www.jmir.org/2021/11/e26522

9. Sampaio RC, Sabbatini M, Limongi R. Diretrizes para o uso ético e responsável da Inteligência Artificial Generativa: um guia prático para pesquisadores. São Paulo: Editora Intercom [Internet]. 2024 [cited 2025 Aug 20]. 62 p. Available from: https://prpg.unicamp.br/noticias/lancamento-diretrizes-para-o-uso-etico-e-responsavel-da-inteligencia-artificial-generativa-um-guia-pratico-para-pesquisadores/

10. Shi J, Chen H, Shi C, Su C, Yue S, Li W, et al. Development and comparative evaluation of knowledge graph-enhanced large language models for domain-specific question answering in nursing. BMC Nursing [Internet]. 2026 [cited 2026 May 5];25:530. Available from: https://doi.org/10.1186/s12912-026-04647-3

11. Gurler N, Sengul T, Bulbul SH, Akyaz DY, Guler O, Yantac AE, et al. Secure and accessible AI in nursing education: a modular agentic chatbot framework comparing ChatGPT-4o with an open-source LLM for chronic wound care. Nurse Educ Pract [Internet]. 2026 [cited 2026 May 5];93:104789. Available from: https://doi.org/10.1016/j.nepr.2026.104789

12. Bsharat SM, Myrzakhan A, Shen Z. Principled instructions are all you need for questioning LLaMA-1/2, GPT-3.5/4. arXiv [Internet]. 2024 [cited 2025 Sep 19];2:1-26. Available from: https://arxiv.org/html/2312.16171v2

13. Pan American Health Organization (PAHO). AI Prompt design for public health: using generative AI responsibly [Internet]. OPAS; 2025. [cited 2025 Sep 30]. 61 p. Available from: https://iris.paho.org/items/b98d9c0f-6008-4b78-9066-01e1e81c29c5

14. Schulhouff S, Llie M, Balepur N, Kahadze K, Liu A, Si C, et al. The prompt report: a systematic survey of prompt engineering techniques. arXiv [Internet]. 2024 [cited 2025 Sep 30];6:1-80. Available from: https://arxiv.org/abs/2406.06608

15. Vilakati, S. Prompt engineering for accurate statistical reasoning with large language models in medical research. Front Artif Intell [Internet]. 2025 [cited 2025 Nov 18];8:1658316. Available from: https://doi.org/10.3389/frai.2025.1658316

16. Dhuliawala S, Komeili M, Xu J, Raileanu R, Li X, Celikyilmaz, et al. Chain-of-verification reduces hallucination in large language models. arXiv [Internet]. 2023 [cited 2025 Oct 12];2:1-19. Available from: https://arxiv.org/abs/2309.11495

17. Bernardes, RM. Prevenção e manejo da lesão por pressão: segurança do paciente [Internet]. São Paulo: Universidade de São Paulo; 2020 [cited 2025 Sep 15]. Available from: http://eerp.usp.br/feridascronicas/recurso_educacional_lp_1_4.html

18. National Pressure Injury Advisory Panel (NPIAP). Prevention and treatment of pressure ulcers/injuries: the international guideline [Internet]. Washington; 2019 [cited 2025 Aug 15]. 405 p. Available from: https://static1.squarespace.com/static/6479484083027f25a6246fcb/t/6553d3440e18d57a550c4e7e/1699992399539/CPG2019edition-digital-Nov2023version.pdf

19. Potter PA, Perry AG, Stockert PS. Fundamentos de enfermagem. 11th. Rio de Janeiro: Guanabara Koogan; 2024. 1644 p.

20. Zhang C, Zhang S, Wu B, Zou K, Chen H. Efficacy of different types of dressings on pressure injuries: systematic review and network meta-analysis. Nurs Open [Internet]. 2023 [cited 2025 Aug 10];10(9):5857-67. Available from: https://doi.org/10.1002/nop2.1867

21. Holman, M. Using tap water compared with normal saline for cleansing wounds in adults: a literature review of the evidence. J Wound Care [Internet]. 2023 [cited 2025 Oct 10];32(8):507-12. Available from: https://doi.org/10.12968/jowc.2023.32.8.507

22. dos Santos TO, da Silva JVL, Rocha RG, Assad LG, da Costa CCP, Pires BMFB. Papain with urea cream in pressure injuries: a case series study. Rev Enferm Atenção Saúde [Internet]. 2024 [cited 2025 Oct 10];13(1):e202404. Available from: https://seer.uftm.edu.br/revistaeletronica/index.php/enfer/article/view/6950

23. O'Connor S, Peltonen LM, Topaz M, Chen LYA, Michalowski M, Ronquillo C, et al. Prompt engineering when using generative AI in nursing education. Nurse Educ Pract [Internet]. 2024 [cited 2025 Nov 19];74:103825. Available from: https://doi.org/10.1016/j.nepr.2023.103825

24. Wang MH, Jiang X, Zeng P, Li X, Chong KKL, Hou G, et al. Balancing accuracy and user satisfaction: the role of prompt engineering in AI-driven healthcare solutions. Front Artif Intell [Internet]. 2025 [cited 2025 Oct 19];8:1517918. Available from: https://doi.org/10.3389/frai.2025.1517918

25. Wang M, Cui J, Lee SMY, Lin Z, Zeng P, Li X, et al. Applied machine learning in intelligent systems: knowledge graph-enhanced ophthalmic contrastive learning with "clinical profile" prompts. Front Artif Intell [Internet]. 2025 [cited 2025 Oct 19];8:1527010. Available from: https://doi.org/10.3389/frai.2025.1527010

26. Flemyng E, Noel-Storr A, Macura B, Gartlehner G, Thomas J, Meerpohl JJ, et al. Position statement on artificial intelligence (AI) use in evidence synthesis across Cochrane, the Campbell Collaboration, JBI, and the collaboration for environmental evidence 2025. JBI Evid Synth [Intermet]. 2025 [cited 2025 Oct 10];23(11):2162-6. Available from: https://doi.org/10.11124/JBIES-25-00480

27. Krifors A, Beskow T, Jonsson M, Lindner KJ, Calås J, Arwsbo V, et al. Improving medication error classification using a reasoning large language model. JAMIA Open [Internet]. 2026 [cited 2025 May 5];9(1):ooag004. Available from: https://doi.org/10.1093/jamiaopen/ooag004

28. Kim SK, Kim GM, Cha S, Jung M. Verification of the validity and reliability of therapeutic communication responses generated by large language models (LLMs): a comparative study of prompt-engineering strategies. J Korean Acad Psychiatr Ment Health Nurs [Internet]. 2025 [cited 2025 May 5];34(Spec):23-35. Available from: https://doi.org/10.12934/jkpmhn.2025.34.S1.23

29. Klein A, Ayoub N, Juhel C, Schuller R, Armstrong F, Pegalajar-Jurado A. Performance of a gelling fibre dressing in management of wounds in a community setting: a sub-analysis of the VIPES study. J Wound Care [Internet]. 2024 [cited 2025 Nov 19];33(7):464-73. Available from: https://doi.org/10.12968/jowc.2024.0125

30. Kimiafar K, Sarbaz M, Tabatabei SM, Ghaddaripouri K, Mousavi AS, Raei M, et al. Artificial intelligence literacy among healthcare professionals and students: a systematic review. Front Health Inform [Internet]. 2023 [cited 2026 May 5];12:168. Available from: https://www.researchgate.net/publication/375600294_Artificial_Intelligence_Literacy_Among_Healthcare_Professionals_and_Students_A_Systematic_Review

Received: 08/12/2025
Approved: 07/05/2026
Associate editor: Dra. Juliana Balbinot Reis Girondi

Corresponding author:
Letícia Pereira Pires
Universidade Federal Fluminense
Rua Recife, Lotes 1-7 - Jardim Bela Vista, Rio das Ostras - RJ, 28895-532
E-mail: leticiapires@id.uff.br

Role of Authors:
Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work - **Pires LP, Goulart MCL, Góes FGB, Ribeiro YC, Guimarães LLC, de Sousa LF.** Drafting the work or revising it critically for important intellectual content - **Pires LP, Goulart MCL, Góes FGB, Ribeiro YC, Guimarães LLC, de Sousa LF.** Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved - **Pires LP, Goulart MCL, Góes FGB, Ribeiro YC, Guimarães LLC, de Sousa LF.** All authors approved the final version of the text.

Conflicts of interest:
The authors have no conflicts of interest to declare.

Data availability:
The authors declare that all data are fully available within the article.

ISSN 2176-9133