CLARITY, SPECIFICITY, AND CONTEXT (CSC): ENGINEERING EFFECTIVE PROMPTS FOR GENERATIVE AI IN HIGHER EDUCATION

Laura Zizka
Learning, EHL Hospitality Business School HES-SO University of Applied Sciences and Arts Western Switzerland (Switzerland)

Abstract

Since 2022, there has been a revolution in Higher Education Institutions (HEIs) called Generative Artificial Intelligence, or Gen AI. Gen AI is based on algorithms that are created to help it learn patterns from large sets of data to produce new content. While other technology, such as the personal computer or the Internet, has been introduced into HEIs, Gen AI has led to heated academic debates and opposed positions regarding its place in HEIs and its use by students and faculty members. For example, some argue that Gen AI has become more trustworthy and reliable, while others critique Gen AI's inability to do critical thinking or complex analytical tasks. ChatGPT and other technologies have been labeled 'disruptive innovations' that could revolutionize academia. Of the many examples of Gen AI, one of the most prevalent is ChatGPT. This study will investigate one of the lesser-researched areas in the research thus far, i.e., prompt engineering. Previous research has defined *prompt engineering* as the strategic means of producing the ideal prompt that will help the user to achieve a specific goal. Gen AI functions by giving a detailed response according to instructions in a prompt. No prompt means no output. Nonetheless, there are best practices to consider when drafting the prompts. According to the literature, effective prompts are based on clarity, specificity, and context. Thus, we drafted ten versions of the same prompt, and ChatGPT drafted ten to gauge if both sets of prompts were equally effective. A content analysis was conducted to evaluate the effectiveness of the prompts. The prompts were also assessed using a rubric based on three criteria – clarity, specificity, and context. All twenty prompts were compared to create a ranking of prompt efficacy. We found that several of our prompts ranked higher than those created by ChatGPT. From the results, our prompt ranked first with a perfect score of 30, although the scores for all prompts were quite close. To conclude, we will share best practices for using prompt engineering in HEIs and tips for writing better prompts.

Keywords: *Prompt engineering, Generative Artificial Intelligence, ChatGPT, Higher Education Institutions (HEIs), academic writing.*

1. Introduction

The history of using technology as a learning tool has followed stages from Web 1.0 (PC Internet) through Web 2.0 (mobile Internet stage) and continuing to Web 3.0 (metaverse and future form of Web) (Wang et al., 2023). In the early stages (before 2010), it was not that popular, but in the development stage, from mid-2010 to 2022, AI algorithms emerged and began to grow rapidly (Wang et al., 2023). In 2022, the technology reached a new high, *flourishing* stage. The use of this technology at each stage has become prevalent in Higher Education Institutions (HEIs). Generative Artificial Intelligence, or Gen AI, is one of the most debated technologies. Gen AI is based on algorithms that are created to help it learn patterns from large sets of data to produce new content (Wang et al., 2023). Open AI was established in 2015 to develop public tools such as ChatGPT; in simple terms, ChatGPT is a highly sophisticated chatbot based on Large Language Models (LLMs) that can answer almost any text-based request as they are based on an exceptional amount of data to train them (Lund et al., 2023). Gen AI functions by giving a detailed response according to instructions in a prompt (Wu et al., 2023).

According to the literature, *prompt engineering* is defined as 'crafting optimal prompt to achieve a specific goal with a generative model" (Amatriain, 2024, p. 3). Effective prompts could include instructions, in-context examples, descriptions of features in the dataset, and questions to encourage fair, unbiased prompts (Li et al., 2024). Heinig (2024) suggested that the most effective prompts are based on clarity (i.e., concise, brief, and easy to understand), specificity (i.e., clear and detailed), and context (i.e., based on relevant background information). Amatriain (2024) suggested that credibility can be

evaluated by prompting LLM to produce multiple answers to the same question. Thus, this study will investigate one of the lesser-researched areas in the research thus far, i.e., prompt engineering through the lens of clarity, specificity, and context (CSC).

2. Methods

To analyze the clarity, specificity, and context of effective prompts, we began by drafting ten versions of the same prompt. We then tasked the paid version of ChatGPT to draft ten prompts to gauge if both sets of prompts were equally effective. A content analysis was conducted to evaluate the prompts' effectiveness and gauge if both the human and the Gen AI-produced prompts were equally effective. Gen AI was then used to create a ranking of prompt efficacy separately and then combined.

Table 1 shows the ten prompts from the researcher and ChatGPT. To save space, only the sentence that differs appears on the table. The full prompt read:

Write a 500-word response to the following position: "This paper argues that hoteliers promote the human-touch, thus, robots cannot replace humans." The response must include 5 academic journal articles used in-text and on a reference list in APA 7th edition format.

Table 1. Ten prompts from the Researcher and ten from ChatGTP.

Researcher ten original versions	ChatGPT ten versions
1." This paper posits that humans cannot be replaced by robots in the hospitality industry, thus, hoteliers should invest more in promoting the human touch. "	1b " This study argues that robots are unable to replicate the human element in the hospitality sector, suggesting that hotels should prioritize investments in enhancing human interactions."
2." This paper argues that humans cannot be replaced by robots in the hospitality industry, thus hoteliers should invest more in people. "	2b " The paper contends that robots lack the capability to replace humans in the hospitality industry, and thus, hoteliers should focus on fostering personalized guest experiences."
3." This paper argues that robots cannot replace humans in the hospitality industry, thus hoteliers should focus on its employees. "	3b. " This analysis proposes that automation cannot fully substitute the human presence in the hospitality field, urging hotel operators to invest in strengthening the personal touch of their services."
4." This paper states that the hospitality industry is people-centric; hence, robots cannot replace humans. "	4b. " This research suggests that robots fall short of replacing human staff in hospitality, advocating for increased investment by hoteliers in elevating the quality of human engagement ."
5." This paper posits that the hospitality industry is human-centric; thus, humans cannot be replaced by robots. "	5b. " The paper asserts that robots cannot replace the unique human skills vital to the hospitality industry, recommending that hotels prioritize efforts to enhance interpersonal customer service."
6." This paper postulates that robots cannot currently replace humans in the hospitality industry as the industry is based on human contact."	6b. " This study maintains that the hospitality industry cannot rely solely on robots to meet guest expectations, and advises hoteliers to invest in human-centric service improvements."
7." This paper posits that human contact is essential in the hospitality industry; thus, robots cannot replace humans ."	7b. " The research argues that while technology can assist, it cannot replace the nuanced human interactions critical to hospitality, prompting hoteliers to focus on promoting the human touch."
8." This paper posits that human contact is essential in the hospitality industry; thus, humans cannot be replaced by robots. "	8b. " The paper posits that robots lack the empathy and adaptability necessary to replace humans in hospitality, urging hotel management to emphasize investments in human-driven service experiences."
9." This paper argues that hoteliers should invest more money in its people because robots cannot replace humans."	9b. " This analysis suggests that the hospitality industry should not view robots as a complete replacement for human staff, and instead, should allocate resources to nurturing the human aspect of guest service."
10." This paper argues that hoteliers promote the human-touch, thus, robots cannot replace humans."	10b. " This paper contends that robots cannot replicate the human warmth essential in hospitality, leading to the conclusion that hoteliers should focus their investments on promoting personal interactions with guests."

Source: OpenAI, 2025.

From Table 1, the length of the examples is the most flagrant and immediate observation. The Researcher's ten prompts resulted in 198 words total compared to 290 words for the ChatGPT output. The Gen AI output added superfluous and overly complicated word constructions that did not contribute to *better* prompts. When asked whose prompts were more effective overall, ChatGPT confirmed that the Researcher's prompts were better. ChatGPT admitted that the variations it provided were too verbose: *My variations use more complex phrasing, such as "This analysis proposes" or "The research argues." While this may seem more academic, it can also introduce ambiguity and make the argument less direct. In academic writing, clarity often outweighs complexity, especially when discussing industry-specific topics* (OpenAI, 2025).

.

426

3. Findings / discussion

The Researcher tasked ChatGPT with the creation of a rubric to rank the outputs: *Create a rubric to evaluate all twenty prompts that focuses on the following three criteria: Clarity, Specificity, and Context. Grade each of the prompts on those three criteria and rank them.*

Rubric for Evaluating Prompts
ChatGPT: *I will create a rubric based on the following three criteria: Clarity, Specificity, and Context. Each criterion will be scored on a scale from 0 to 10, making the total score out of 30 for each prompt.*

Table 2. Rubric for evaluating prompts.

Clarity	Specificity	Context
Definition: Measures how clearly the prompt conveys its main argument. A clear prompt has a straightforward thesis and avoids vague or complex language.	Definition: Evaluates the level of detail and precision in the prompt. A specific prompt clearly defines its terms and offers a precise recommendation or action.	Definition: Assesses how well the prompt addresses the unique characteristics of the hospitality industry, including the importance of human touch and the role of automation.
Scoring: □9-10 points: Thesis is crystal clear, concise, and easily understandable. □7-8 points: Mostly clear, but may include minor wording issues or slight ambiguity. □5-6 points: Somewhat clear, but has noticeable issues with vague language or complex phrasing. □0-4 points: Confusing or unclear thesis; difficult to discern the main argument.	Scoring: 9-10 points: Provides a well-defined argument with specific recommendations or detailed language. 7-8 points: Generally specific, but may lack detailed action points or use slightly vague language. 5-6 points: Lacks precision or uses generic language without specific recommendations. 0-4 points: Very vague, with broad statements that lack any detail or actionable insight.	Scoring: □9-10 points: Strongly aligned with the specific context of hospitality, addressing industry-specific issues effectively. □7-8 points: Good contextual relevance, but could more deeply integrate industry-specific aspects. □5-6 points: Shows some understanding of the context but may be too general or lack specific industry focus. □0-4 points: Weak or no connection to the hospitality context; could apply to any industry without significant modification.

Source: OpenAI, 2025.

Table 2 shows the specific criteria created by ChatGPT to evaluate the prompts. ChatGPT's definition of *clarity* focuses on 'straightforward, direct language,' which mirrors Heinig's (2024) criteria of 'concise, brief, and easy to understand.' *Specificity* for ChatGPT includes 'detail' and 'precision,' which concurs with Heinig's 'clear and detailed criteria (2024). Finally, ChatGPT's definition confirms Heinig's (2024) proposition of providing relevant background. In fact, ChatGPT elaborates on the industry and links strongly to the prompt when mentioning the human touch. Thus, ChatGPT can create criteria similar to a human for the definition. However, the point allocation for scoring is a bit fuzzy. What would differentiate a seven from an eight if the criteria were applied? To mitigate that issue, a human faculty member could rework the scoring to be more user-friendly for their grading.

The Researcher then requested ChatGPT to assess all twenty prompts using this rubric.
The ranking of the 20 prompts goes as follows:

Table 3. Grading of the 20 prompts (together).

Grading of All 20 Prompts

Prompt	Clarity (0-10)	Specificity (0-10)	Context (0-10)	Total Score (0-30)
Your Variation 1	10	10	10	30
My Variation 1	10	9	10	29
Your Variation 5	9	9	10	28
My Variation 5	9	9	9	27
Your Variation 7	9	8	10	27
My Variation 3	8	9	9	26
My Variation 7	8	8	9	25
Your Variation 2	8	7	9	24
My Variation 10	8	8	8	24
Your Variation 8	8	7	8	23
My Variation 2	7	7	9	23
Your Variation 3	8	6	8	22
My Variation 4	7	7	8	22
Your Variation 6	6	7	7	20
My Variation 9	6	6	8	20
Your Variation 10	5	6	7	18
My Variation 8	6	5	7	18
Your Variation 4	5	5	7	17
Your Variation 9	5	5	6	16
My Variation 6	5	4	6	15

Key: Your variation= Researcher   My variation: ChatGPT
Source: OpenAI, 2025

While it was not necessarily expected, the versions from the Researcher and ChatGPT both had five examples in the top ten. Nonetheless, the Researcher's versions would, overall, score more points. When asked to explain this phenomena, ChatGPT responded that the Research versions were clear and direct, consistent and focused, and relevant to the industry (OpenAI, 2025). ChatGPT's more detailed response was the following:

1. Clarity and Directness:
- Your variations use clear, simple, and direct language. For instance, "This paper posits that humans cannot be replaced by robots in the hospitality industry" is straightforward and easy to understand.
- The focus on terms like "human contact," "people-centric," and "human touch" provides a strong emphasis on the core values of the hospitality industry without ambiguity.

2. Consistency and Focus:
- Your variations are consistent in their emphasis on the central theme: the irreplaceable value of human employees in hospitality. There is less deviation from this main point, making the argument cohesive.
- Many of your prompts explicitly highlight the industry's reliance on human interactions and the need for investment in the human element (e.g., "invest more in promoting the human touch"). This consistency would likely be appreciated by graders for maintaining a clear, focused argument throughout.

3. Relevance to the Industry:
- Your prompts use industry-specific language effectively, such as "people-centric," "human contact," and "human touch." These phrases directly address the unique characteristics of the hospitality industry, demonstrating an understanding of what differentiates it from other sectors.
- The focus on investment in people and the human element aligns well with industry trends, making your variations more relevant and applicable to real-world contexts.

Source: OpenAI, 2025.

ChatGPT further elaborated on why its scores would be lower, citing complex phrasing which could lead to greater ambiguity, less consistency in actionable recommendations that could weaken them, and a lack of logical flow which reduces the effect of the prompts (OpenAI, 2025).

Further, when asked, ChatGPT provided a critique of each of the 20 prompts as seen on Table 4.

Table 4. Comments on the 20 prompts.

Final Comprehensive Ranking of All 20 Variations

Rank	Prompt	Score	Reason for Ranking
1	Your Variation 1	10/10	Clear thesis, strong recommendation, direct language, and industry relevance.
2	My Variation 1	9.5/10	Well-articulated argument, slightly less direct than the top prompt.
3	Your Variation 5	9/10	Emphasizes the human-centric nature of the industry effectively.
4	My Variation 5	8.5/10	Focuses on unique human skills and provides a clear call to action.
5	Your Variation 7	8/10	Highlights the essential role of human contact; straightforward and relevant.
6	My Variation 3	7.5/10	Solid argument, but slightly wordy. Good actionable recommendation.
7	My Variation 7	7.5/10	Balanced approach, acknowledging technology's role while emphasizing human interactions.
8	Your Variation 2	7/10	Simple and direct, but slightly vague in the recommendation ("invest more in people").
9	My Variation 10	7/10	Strong focus on personal interactions but could be clearer in language.
10	Your Variation 8	6.5/10	Similar to Your Variation 7 but slightly redundant in phrasing.
11	My Variation 2	6.5/10	Argument is clear but lacks specificity in the recommendation.
12	Your Variation 3	6.5/10	Clear argument but could benefit from more specific language about employee focus.
13	My Variation 4	6/10	Emphasizes quality of engagement but is slightly vague in the call to action.
14	Your Variation 6	6/10	Weakened by the use of "currently," implying robots might replace humans in the future.
15	My Variation 9	5.5/10	Redundant language and vague recommendation. Lacks a strong, direct stance.
16	Your Variation 10	5/10	Weak logical flow and unclear connection between hoteliers promoting the human touch and robots' inability to replace humans.
17	Your Variation 4	5/10	Simplistic argument, lacking depth and actionable insight.
18	My Variation 8	5/10	Argument is vague, lacking strong evidence or specific recommendations.
19	Your Variation 9	4.5/10	The language is vague and does not clearly specify how hoteliers should invest in people.
20	My Variation 6	4/10	Lacks assertiveness and depth, focusing too much on general service improvements without specific recommendations.

Source: OpenAI, 2025.

ChatGPT's comments on its reasons for ranking merit further analysis and reflection. With an equal number of highly ranked prompts from the Researcher and ChatGPT, it was necessary to try to gauge why one prompt was 'better' than another. While the Researcher's prompts were criticized for language, weak flow, and simplicity, ChatGPT's versions had weaker arguments, call to action, and recommendations.

4. Conclusions / limitations / future studies

Based on the results of this study, a writer may wonder why ChatGPT would not provide 'perfect' examples each time; in fact, the question remains: why did ChatGPT not provide better examples than a human? One response derives from the method in which Gen AI produces output: For example, ChatGPT uses a simple information processing method, i.e., it predicts the next possible word in a sentence (Wu et al., 2023). There is no magical, technology-writing genius behind the screen. The technology does not think; it does not reflect; it does not try to help one writer over another. Thus, some writing output from GenAI may produce harmful or biased content (Tiulkanov, 2023) because the technology cannot think (Li et al., 2024). In fact, output may not be truthful because it cannot care if it is truthful or not and is not predictable about how or when it lies (Tiulkanov, 2023). Moving forward, to create more effective prompts, users may choose to focus on Heinig's (2024) three criteria: Clarity, Specificity, and Context.

This study was limited in its depth. Twenty prompts may not be sufficient to draw profound conclusions. Nonetheless, this gave a first glimpse into the efficacy of prompt engineering. Thus, a future study could elaborate on this initial study by drafting hundreds of examples by one or several researchers. The ranking was done uniquely by ChatGPT. A further study could include the Researcher and ChatGPT both ranking the prompts to compare their responses. Finally, ChatGPT was the only Gen AI used in this study. While it is free and easily accessible to all (which is why it was chosen), many other Gen AI could have been used. A study could be conducted using several Gen AI and comparing how they differ in their output and results.

References

Amatriain, X. (2024). Prompt design and engineering: Introduction and advanced methods. *arXiv:2401.14423*. https://doi.org/10.48550/arXiv.2401.14423

Heinig, I. (2024). *How to write the best AI prompts (+examples, do's and don'ts).* Retrieved January 8, 2025, from: https://sendbird.com/blog/ai-prompts

Li, Y., Zhang, L., & Zhang, Y. (2024). Fairness of ChatGPT. *arXiv:2305.18569*. https://doi.org/10.48550/arXiv.2305.18569

Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimary, S., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of Association for Information Science and Technology, 74*(5). https://doi.org/10.1002/asi.24750

OpenAI. (2025). *ChatGPT (October 25 version)* [Large language model]. https://chatgpt.com/c/671b78f0-d744-800e-8692-b44fbaa78911

Tiulkanov, A. (2023, January 8). Is it high time to take ChatGPT offline? *AI, Data & Digital Policy*. Retrieved from https://www.linkedin.com/pulse/high-time-take-chatgpt-offline-aleksandr-tiulkanov

Wang, Y., Pan, Y., Yan, M., Su, Z., & Luan, T. H. (2023). A survey on ChatGPT: AI-generated contents, challenges, and solutions. *IEEE Open Journal of the Computer Society*, *4*, 280-302. doi: 10.1109/OJCS.2023.3300321

Wu, T., He, S., Liu, J., Sun, S., Liu, K., & Han, Q.-L. (2023). A brief overview of ChatGPT: The history, status quo and potential future development. *Journal of Automatica Sinaca, 10*(5), 1122-1136. doi: 10.1109/JAS.2023.123618