Human Systems Engineering and Design (IHSED 2026), Vol. 237, 2026, 210–219
https://doi.org/10.54941/ahfe1008161

Evaluating Prompting Strategies for Spatial Planning in Mixed-Use Architecture: A Case Study of Kalkbreite, Zürich

Seyed Damoun Pezeshki, Ali Ghazvinian, and Mahyar Hadighi

Texas Tech, Lubbock, TX 79409, USA

ABSTRACT

In this paper, we will explore how different prompts affect the performance of LLMs when it comes to creating spatial plans in architecture. There are four strategies used in this study: one-shot prompting, few-shot prompting, chain-of-thought prompting, and meta-prompting. These methods are tested by presenting the same example of a mixed-use building, which is graded according to efficiency, complexity, accuracy, and effort involved in each case. It has been found that all these methods can provide a reasonably accurate output, although the quality and efficiency vary. Meta-prompting and chain-of-thought prompting are found to be quite detailed, while one-shot prompting is quick yet lacks detail. Few-shot and chain-of-thought prompting enable users to control their outputs through an interactive process, but this method takes up considerable time and effort.

Keywords: Artificial intelligence, Prompt engineering, Computational design, Spatial planning

INTRODUCTION

With the fast-paced evolution of LLMs, the possibilities of their employment in architecture have increased as well. In particular, the use of such large language models becomes especially relevant when it comes to dealing with the issues of spatial programming, planning, organizing and reasoning related to the architectural practice. The development of AI's ability to process sophisticated inputs and provide structured outputs has created a situation where the process of interacting with AI through prompting became a significant determinant of the results' quality and reliability.

As the process of prompting evolves, new approaches to instructing the artificial intelligence on what needs to be done emerge. For instance, prompting can vary in terms of its complexity and efficiency depending on whether a technique like chain-of-thought prompting, meta-prompting or even one-shot prompting is applied. Each method implies using a certain level of structuring and cognitive guidance that affects not only the interaction's effectiveness, but also the quality of the output itself. Even though previous studies have explored such techniques in relation to general-purpose AI tasks, their analysis in an architectural spatial planning context is rather scarce.

Received May 8, 2026 Revised June 20, 2026 Accepted July 3, 2026 Available online September 25, 2026

Such knowledge is essential for filling the gap that exists today, and to achieve this purpose, the proposed research will consider multiple prompt techniques within the context of the architecture. Using an example of a complicated mixed use building, the study will focus on evaluating the effect of various prompting procedures on the AI's ability to process the information provided, recognize the need for certain spaces, and offer solutions.

As part of the work, the research will create an evaluation matrix, based on which the effectiveness, accuracy, detail, and difficulty of prompts will be analyzed. It will become possible to identify not only the most effective prompt technique, but also the conditions for its optimal application. Such knowledge will make it easier to integrate artificial intelligence into architects' daily routine.

DIFFERENT STRATEGY OF PROMPTING

Prompt is the primary input to an AI model, typically a piece of text, that includes instructions, context, and examples to elicit a specific response. (Giray, 2022). Prompting involves various strategic approaches. Zero-shot prompting involves providing an AI model with a task or query without any explicit examples in the prompt itself (Jyoti et al., 2025; Yurchak et al., 2024). The model relies solely on its pre-trained knowledge to generate a response (Jyoti et al., 2025; Basile, 2021) One-shot prompting is an extension of zero-shot where a single example is provided within the prompt to guide the model's understanding of the task (Jyoti et al., 2025). Few-shot prompting involves incorporating a small number of demonstration examples directly into the prompt (Jyoti et al., 2025; Bahrami et al., 2023). These examples help the model understand the task's requirements and desired output format by providing context and patterns (Bahrami et al., 2023). Instruction-based prompting uses explicit natural language directives within the prompt to describe the desired task, constraints, or format (Giray, 2022; Kovari, 2024; Viswanathan, 2025). Chain-of-Thought (CoT) prompting encourages the model to generate a sequence of intermediate reasoning steps before arriving at a final answer (Sahoo et al., 2023; Qiao et al., 2022). This approach mimics human thought processes and makes the AI's reasoning explicit (Qiao et al., 2022). Prompt chaining involves decomposing a complex task into a series of smaller, sequential subtasks, each addressed by a separate prompt that builds upon the output of the previous one (Wu et al., 2021; Viswanathan, 2025). Meta-prompting is a sophisticated paradigm where LLMs are prompted to recursively generate and refine prompts themselves, often grounded in formal theories like type and category theory (Zhang et al., 2023; Wynter et al., 2023). This allows the AI to develop better prompts for a given task (Zhang et al., 2023; Ye et al., 2023). Role-play prompting instructs the AI to assume a specific persona, character, or role to influence its responses and reasoning (Brown et al., 2024; Saputra et al., 2024; Mzwri & Turcsányi-Szabo, 2024).

Figure 1: Picture of Kalkbreite, Zürich (Google maps satellite view, 2005).

Methodology

In this research work, the comparative experiment method will be used in evaluating the effect that different types of prompts have on the functioning of LLMs within the framework of architectural space planning activities. The case study will be a single one called Kalkbreite in Zürich, which has been chosen because of its complex program and combination of residential, commercial, and public purposes.

The methodology begins with the identification and selection of four distinct prompting strategies: one-shot prompting, few-shot prompting, chain-of-thought prompting, and meta-prompting. These strategies represent varying levels of instructional depth and cognitive guidance provided to the AI model.

Experimental procedure consists of a well-planned process with an iterative cycle for each task. The process involves designing prompts using the selected strategy and presenting them to the LLM. After evaluating their outputs, where needed, they iteratively adjust the prompts until generating a usable plan. All of the processes and steps taken during the experiment are systematically recorded.

In order to facilitate a comparative analysis, the effectiveness of each approach is determined through the use of a multiple-criteria grading system. The grading system has been designed to evaluate efficiency, accuracy, and reliability through objective criteria.

Efficiency is evaluated using three measurable indicators: (1) the number of words per prompt, (2) the number of iterations required to reach a usable output, and (3) the length of reasoning reflected in the response. Each indicator is scored on a scaled system from 1 to 5, where lower resource consumption and fewer iterations correspond to higher scores. For example, prompts requiring fewer words and iterations receive higher efficiency scores, while excessively long or repetitive prompting processes receive lower scores.

Accuracy and design relevance are assessed qualitatively but translated into numerical values using a standardized rubric. Outputs are evaluated based on program coherence, spatial logic, and consistency with both the documented characteristics of the case study and established mixed-use design principles. Each of these criteria is also scored on a 1 to 5 scale, where a score of 5 indicates a highly coherent and realistic spatial proposal, and a score of 1 indicates significant conceptual or organizational flaws.

To address the presence of misinformation or unrealistic outputs, a penalty-based reliability assessment is incorporated into the grading system. Outputs are reviewed for major factual inaccuracies, spatial inconsistencies, or non-viable design suggestions. When such issues are identified, negative points (–1 to –3) are assigned depending on the severity of the error. Minor inconsistencies result in small deductions, while critical misunderstandings of spatial logic or program result in larger penalties.

This final score will be calculated using a sum of efficiency and accuracy scores for each prompting technique, with reliability penalties being applied to it when necessary. Thus, we obtain an overall score for the technique based on its efficiency and the validity of its implementation. The grading system prevents us from valuing a high efficiency yet low-quality result, ensuring that only good solutions are rewarded.

The comparative analysis that follows scoring involves examining trends, advantages, and weaknesses inherent in the prompting strategies. The comparison will be done with regard to efficiency vs. accuracy and applicability of each strategy to different kinds of spatial planning scenarios. Unlike the use of complicated statistical analysis models in similar research works, our study utilizes an easy-to-understand comparative method of analysis.

The results of this approach include creating a decision-making model for the determination of the best prompting techniques for different architectural projects. The creation of a formalized grading system helps ensure transparency and replicability in the studies involving architectural AI in the future.

Introducing the Building

The building is a large, multi-level structure organized over eight floors (including a basement) with a total area of 23,575 m², suggesting a complex institutional or mixed-use function. Its massing expands at the base and gradually tapers upward, creating a grounded, monumental lower level and lighter, more articulated upper floors. The ground floor, the largest, functions as the main public platform and contains a prominent triple-height space that extends into the first and second floors, reducing their footprint. Circulation is defined by a dual-access system: a primary entrance at ground level leading into a large atrium, and a secondary via an external stair to a third-floor courtyard. This courtyard acts as a key transitional and social space, linking vertically to upper levels. Overall, the building establishes a clear spatial hierarchy, with open, functions at the base and more private or specialized areas above, creating a dynamic sequence of movement and spatial experience.

The introduction of the building to the AI for creating the description of it is presented in this section. To make sure that the input was framed consistently, the prompt that was used for the purpose was created using a different AI agent. With this kind of approach, it becomes possible to achieve neutrality regarding the creation of the input, thus eliminating potential biases that may appear in the course of research carried out by a person. The amount and kind of information about the building were consistent across all the types of prompting. Still, while one-shot and few-shot prompting resulted in a representation of the building similar to how human beings would describe it, the meta-prompting approach produced a slightly modified input based on human description but created in a specific style to maximize its interpretability for the AI.

Table 1: Introducing the building.

Prompt Strategy	Number of Words (Input)
One-shot prompting	110 of 321 (Belong to this part)
Few-shot prompting	110
Meta-prompting	356 of 1312 (Belong to this part)
Chain-of-thought prompting	114

INTRODUCING NEEDS

In the "Introducing Needs" phase, a similar methodology to the previous section is applied. The prompt used to define the spatial and functional requirements of the building is generated by a separate AI agent and then provided to the primary AI responsible for spatial planning. This ensures a consistent and unbiased articulation of programmatic needs, including required spaces, functional uses, and performance expectations for the final output. As in the prior stage, the same dataset and level of detail - not same in format- are maintained across all prompting strategies to ensure comparability. For the one-shot and meta-prompt approaches, all relevant information—including building description, spatial requirements, and expected outcomes—is consolidated into a single, comprehensive prompt provided to the AI. In contrast, for the remaining strategies, this phase is divided into two sequential prompts: the first introduces the building, and the second specifies the spatial needs and expected results. This structured variation allows for evaluating how the segmentation or consolidation of input information influences the effectiveness of AI-driven spatial planning.

It is important to mention that the definition of spatial needs and programming requirements at this stage does not involve speculation but is based on the conditions in which the building currently exists. The definition of spaces, their functions, and other information is based on items present in the building itself, thus making sure that the input will be based on realistic and contextual information only. By doing this, it is possible to limit the involvement of any external factors that could lead to inconsistencies in the definition of both the building and the needs. This way, the AI works with data that is quite close to the reality of the building, and it becomes possible to analyse the impact of different prompts on it.

Table 2: Introducing needs.

Prompt Strategy	Number of Words (Input)		Number of Prompt
One-shot prompting	211 of 321 (Belong to this part)		1
Few-shot prompting	133	138	2
Meta-prompting	956 of 1312 (Belong to this part)		1
Chain-of-thought prompting	478	491	2

Discussion

The results from the different prompting techniques showed a high degree of similarity in terms of spatial organization and the meaning that was assigned to the buildings in question. Overall, the AI came up with similar planning outcomes regardless of the type of prompts used. This implies that the core dataset played a more significant role in the planning outcome than the prompting technique. Nonetheless, the degree of comprehensiveness and clarity in the output differed significantly from one prompting method to another. The chain-of-thought and meta-prompting techniques provided the most comprehensive answers to the planning tasks. The AI was able to provide detailed explanations and articulate the reasons behind its spatial arrangements effectively. Moreover, the few-shot and chain-of-thought prompting methods yielded step-by-step responses between prompts. For instance, the AI would generate a response after being prompted about the building characteristics. Consequently, the generated output gave clear indication of whether the AI had correctly understood the dataset used.

Table 3: Total input and output.

Prompt Strategy	Total Input	Total Output	Total Prompt Number
One-shot prompting	321	633	1
Few-shot prompting	381	1164	3
Meta-prompting	1312	998	1
Chain-of-thought prompting	1082	1347	3

Furthermore, one very important thing that should be considered when examining the efficiency of using AI in this research is the efficiency with regard to time and effort taken in order to create the result. From the information provided in Table 3, a certain compromise can clearly be observed regarding the correlation between the degree of input effort and quality of output. It can be seen that the meta-prompt and chain-of-thought techniques are much more labor-intensive because they include either more complex instruction or several stages of interaction. At the same time, such approach leads to much more polished and elaborate output with higher details.

Moreover, the sequential nature of few-shot prompting and chain-of-thought makes it possible to gain better control over the entire procedure. Since the prompt is split into different steps, each stage can be checked beforehand, which means higher confidence in the result. But such better control entails more prompts, which means that the effort put in by the user increases.

In order to measure these differences in an objective way, it is necessary to make a quantitative evaluation of the effectiveness of these approaches. The definition of effectiveness used in this study implies comparing output and input, expressed as word count. It is determined through the following formula:

$$\text{Efficiency} = \frac{\text{Total Output Words}}{\text{Total Input Words}}$$

This metric provides a standardized basis for comparing prompting strategies, enabling a clearer assessment of how effectively each method converts input effort into usable output.

Table 4: Efficiency.

Prompt Strategy	Amount of Efficiency
One-shot prompting	1.972
Few-shot prompting	3.055
Meta-prompting	0.761
Chain-of-thought prompting	1.245

Another important parameter in assessing the results of this study is their accuracy and level of detail. According to the analysis of the results, all methods generated accurate results, and there were no examples of the presence of any falsehoods or distortions. It means that the quality of translation of the building description is consistently high and depends on neither the specific prompting strategy applied nor the consistency of the input data. In such a way, all the methods can be evaluated equally in terms of accuracy, so they receive the same grade for the level of accuracy of the results. The main difference in this case is the degree of detail of the interpretation of the results and the explanation provided. When discussing this particular aspect, one can see that while accuracy stays the same, chain-of-thought and meta-prompting generate more elaborate and detailed results than one-shot and few-shot methods do.

Table 5: Specific output.

Prompt Strategy	Number of Words (Specific Output)
One-shot prompting	147
Few-shot prompting	288
Meta-prompting	384
Chain-of-thought prompting	321

As described earlier, the grading scale used to evaluate prompting strategies uses numbers for comparison purposes. While most criteria are measured using a range of 1 to 5 points, where the higher value represents a more successful performance, the criterion of accuracy differs in that it uses a negative scale, ranging from –3 to 0 points. Here, a positive number indicates the existence and degree of errors in the data, while 0 is used to denote full accuracy. Since there were no notable inaccuracies among the output responses, all prompting strategies achieve a score of 0 in terms of accuracy. Therefore, accuracy does not help differentiate between various approaches, but rather serves to provide a standardized benchmark for reliability. In order to determine which prompting strategy is the best performing, the average score obtained in all criteria for each technique was calculated.

Table 6: Grading.

Prompt Strategy	Effort	Efficiency	Accuracy	Specific Output	Final
One-shot prompting	5	3.2	0	1.9	3.36
Few-shot prompting	1.7	5	0	3.7	3.46
Meta-prompting	5	1.2	0	5	3.73
Chain-of-thought prompting	1.7	2	0	4.1	2.60

*Effort is based on number of prompt that given to AI, the accuracy doesn't have any way in this calculation and the other variable same weight in final score.

CONCLUSION

Based on the analysis of the findings from the conducted research, the meta-prompt strategy appears to be the best option for all the proposed strategies. The meta-prompt strategy produces the most elaborate, coherent, and explained output without losing accuracy and reliability. Therefore, it is the most suitable technique when performing complex tasks, such as spatial planning. Still, the study findings suggest that there is no single strategy that can serve all possible circumstances, and the selection of a specific strategy depends on the particular task needs.

When prompt efficiency and speed are prioritized, but the level of output details does not matter much, the one-shot technique becomes highly efficient. In contrast, the few-shot and chain-of-thought strategies provide an opportunity to regulate the process because they are based on step-by-step operations. Thus, it is possible to control the AI's thinking process. However, the downside is that multiple prompts have to be produced, which takes time and effort.

Despite being very effective in the terms of output quality, the entire process of utilizing the chain-of-thought prompting may take a lot of time due to the numerous interactions involved and the intellectual strain on the part of the operator. At the same time, the use of an approach combining several different approaches may be more productive.

To put it another way, the application of a meta-prompt to generate the first result and its further correction or supplementation through a few iterations of the chain-of-thought prompting would allow achieving the best results in regard to the quality of output while also providing greater precision and flexibility in the process.

In summary, it may be said that the most effective approach in case of communication with AI in architecture would be that of meta-prompting, while the fastest technique would be the one-shot prompting.

ACKNOWLEDGMENT

It is my pleasure to acknowledge my appreciation to Dr. Gazvinian and Dr. Hadighi for their significant contributions to this paper. They have provided me with immense assistance, which helped shape the research into what it is today.

REFERENCES

Bahrami, M., et al. (2023). Few-shot learning with prompting methods. In 2023 6th International Conference on Pattern Recognition and Image Analysis (IPRIA). https://ieeexplore.ieee.org/document/10147172/

Basile, V. (2021). Is EVALITA done? On the impact of prompting on the Italian NLP evaluation campaign. https://www.semanticscholar.org/paper/e10c2ed8de5568e6a41fb60c2adc4bf9c225da25

Becker, A., Kienbaum, L., AA Projects, & Cachola Schmal, P. (Eds.). (2015). Bauen und Wohnen in Gemeinschaft / Building and living in communities: Ideen, Prozesse, Architektur / Ideas, processes, architecture. Birkhäuser. https://www.muellersigrist.ch/assets/Uploads/publications/060/060-Bauen-und-Wohnen-in-Gemeinschaft-Publikation-DAM-Sept2015.pdf

Brown, K., et al. (2024). The impact of prompt phrasing on AI-based chatbot effectiveness and user experience. https://lup.lub.lu.se/student-papers/search/publication/9201564

Giray, L. (2022). Prompt engineering with ChatGPT: A guide for academic writers. Annals of Biomedical Engineering. https://link.springer.com/article/10.1007/s10439-023-03272-4

Jyoti, S., et al. (2025). Optimizing generative AI applications: A comparative study of effective prompting techniques. In 2025 5th International Conference on Pervasive Computing and Social Networking (ICPCSN). https://ieeexplore.ieee.org/document/11035207/

Kovari, A. (2024). ChatGPT the omniscient? A guide to effective prompting. In 2024 IEEE 7th International Conference and Workshop Óbuda on Electrical and Power Engineering (CANDO-EPE). https://ieeexplore.ieee.org/document/10772984/

Mzwri, K., & Turcsányi-Szabo, M. (2024). The impact of prompt engineering and a generative AI-driven tool on autonomous learning: A case study. Education Sciences. https://www.mdpi.com/2227-7102/15/2/199

Qiao, S., et al. (2022). Reasoning with language model prompting: A survey. arXiv. https://arxiv.org/abs/2212.09597

Sahoo, P., et al. (2023). A systematic survey of prompt engineering in large language models: Techniques and applications. https://rotmandigital.ca/wp-content/uploads/2024/09/A-Systematic-Survey-of-Prompt-Engineering-in-Large-Language-Models.pdf

Saputra, Y. S. A., et al. (2024). Advanced prompting techniques for artificial intelligence-based learning innovation. JEPTIRA. https://aslabunsada.my.id/ojs/index.php/jeptira/article/view/98

Viswanathan, P. S. (2025). Prompt engineering for conversational AI systems: A systematic review of techniques and applications. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. https://ijsrcseit.com/index.php/home/article/view/CSEIT25111276

Wu, T., et al. (2021). AI chains: Transparent and controllable human-AI interaction by chaining large language model prompts. In Proceedings of the ACM. https://dl.acm.org/doi/abs/10.1145/3491102.3517582

Wynter, A. de, et al. (2023). On meta-prompting. arXiv. https://arxiv.org/abs/2312.06562

Ye, Q., et al. (2023). Prompt engineering a prompt engineer. In Annual Meeting of the Association for Computational Linguistics. https://arxiv.org/abs/2311.05661

Yurchak, I., et al. (2024). Prompting techniques for enhancing the use of large language models. Computer Systems and Network. https://science.lpnu.ua/csn/all-volumes-and-issues/volume-6-number-2-2024/prompting-techniques-enhancing-use-large-language

Zhang, Y., et al. (2023). Meta prompting for AI systems. https://www.semanticscholar.org/paper/f502cf5c371f3768362816e2fefca1e8d2751341