Use of Language to Generate Architectural Scenery with AI-Powered Tools

Karahan, Hanım Gülsüm; Aktaş, Begüm; Bingöl, Cemal Koray

DOI
10.1007/978-3-031-37189-9_6

Publication date
2023

Document Version
Final published version

Published in
Computer-Aided Architectural Design. INTERCONNECTIONS

Citation (APA)
Karahan, H. G., Aktaş, B., & Bingöl, C. K. (2023). Use of Language to Generate Architectural Scenery with AI-Powered Tools. In M. Turrin, C. Andriotis, & A. Rafiee (Eds.), *Computer-Aided Architectural Design. INTERCONNECTIONS: Co-computing Beyond Boundaries - 20th International Conference, CAAD Futures 2023, Selected Papers* (pp. 83-96). (Communications in Computer and Information Science; Vol. 1819 CCIS). Springer. https://doi.org/10.1007/978-3-031-37189-9_6

Use of Language to Generate Architectural Scenery with AI-Powered Tools

Hanım Gülsüm Karahan[1(✉)], Begüm Aktaş[1,2], and Cemal Koray Bingöl[3]

[1] Architectural Design Computing Graduate Program, Istanbul Technical University, 34367 Istanbul, Turkey
{karahan21,aktas15}@itu.edu.tr

[2] Department of Architecture, Altınbaş University, Istanbul, Turkey

[3] Department of Management in the Built Environment, Delft University of Technology, Delft, The Netherlands
c.k.bingol@tudelft.nl

Abstract. The quality of communication with a computer impacts how the designer performs during the design process. Today, Artificial Intelligence (AI) empowers the designer by expanding the solution space using the expertise from previous knowledge. However, the developments in AI-powered design tools mainly focus on visual and spatial enhancements. In the last decade, AI-powered design tools mostly experimented with image transformation models (GANs) to provide fast insights to designers using learned experiences, simulations, or datasets. The studies on the design process using verbal language with the help of AI are limited. Therefore, designers' capacity to communicate with intelligent machines would lead us to envision the future of AI-powered design tools.

In design practice, designers develop individual and contextual studies through digital tools. This study investigates the process of architectural visual generation and verbal communication to describe architectural images by architecture graduates with prior experience or no experience in prior with Midjourney. The research focuses on the designers' semantic language during the design process with the AI-powered tool and analysis of the verbal part of the communication. The results of this study show that participants' first impressions of the image and how they express their impressions through description do not correspond with how Midjourney interprets those descriptions. Furthermore, architects' image generation process using the tool is nonlinear. As architects develop a deeper understanding of changing modes of interactions, they are more likely to benefit from AI-powered tools as collaborative entities.

Keywords: Artificial Intelligence · Design Cognition · Human-Machine Interaction · Digital Design

1 Introduction

Designing is a complex and temporal activity that requires generating, transforming, and refining images of different aspects of that still non-existent artifact and making representations of it, enabling communication and examination of the ideas involved

M. Turrin et al. (Eds.): CAAD Futures 2023, CCIS 1819, pp. 83–96, 2023.
https://doi.org/10.1007/978-3-031-37189-9_6

[1, 2]. In the design process, a sketch is a reflection of the guiding mental image, but it cannot be identical to it, and this difference is precisely what makes it a precious instrument for the designer. By making a sketch, the designer supplies the mental image with the assistance of an optical image, which has all the properties of such visual perception [3]. The interaction of arguments in the design process, sketch represents the visual perception and exploration process of finding solutions to the design problem and reasoning about it. Working in some visual medium-drawing, in Schön's examples-the designer sees what is "there" in some representation of a site, draws in relation to it, and sees what he or she has drawn, thereby informing further designing [4]. Therefore, seeing triggers new ideas constantly and changes things accordingly to what has been drawn on paper. This conversational structure of seeing-moving-seeing represents an iterative process where every move feeds the next moves or vice versa to construct new meanings.

The discovery of unintentional ideas in this process comprises the dual nature of visual thinking. Visual thinking has the power to reveal both intended and unintended ideas for the design process. The interaction of arguments of "see-move-see" is a representation process of the goal image and realization of the idea in the context. In all this "seeing," the designer visually registers information and constructs its meaning, identifies patterns, and gives them meaning beyond themselves. Words like "recognize," "detect," "discover," and "appreciate" denote variants of seeing, as do such terms as "seeing that," "seeing as," and "seeing in.“ This process of seeing-drawing-seeing" is one example of what Schön means by designing as a reflective conversation with the materials of a situation [4]. Therefore, visual thinking can coalesce abstract and perceptible ideas as one or analyze them separately. In addition, in order to see, we had to think, and we had nothing to think about if we were not looking [5]. It is now well established from various studies that sketching is not just a representation of an idea; but a process of seeing, visual reasoning, and imagination. Sketching is not merely an act of representing a preformulated image; in this context, we often deal with a search for such an image [2]. Also, what the viewer "sees" in the picture is already the outcome of that organizational process [5]. In this process, in the interaction of arguments, sketching reveals the design knowledge of a designer.

2 Reasoning Through Visual and Language

When reflecting on the nature of thinking, most people associate it, primarily with words, with language when visual thinking is considered; however, we tend to concentrate on visual and almost forget thinking, which fades into the background [6]. Physical actions refer to drawing and looking, but yet. Perceptual actions refer to the interpretation of visual information [7], where interpretation is associated with thinking and language. Descriptionalists believe that mental images represent the mode of language rather than pictures [2]. Thinking is mostly associated with language to identify how it is produced and develops over time. Language has syntactic, semantic, and pragmatic rules to serve different purposes of humans or machines. We use language to conceptualize reality in the physical world, and our thinking is directly related to how we perceive the physical world through language. Figure 1 shows us a sentence's mental tree consisting of rules

[8]. This structure can grow with symbols such as "if, then," which explains a reasonable story.

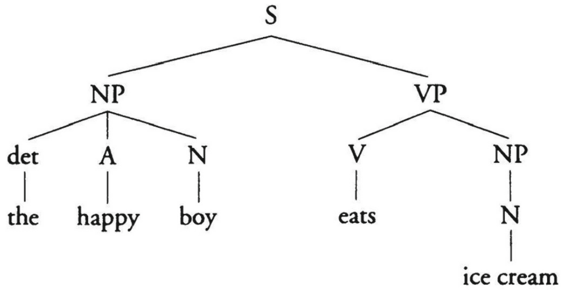


Fig. 1. A mental tree of a sentence (S: sentence, NP: noun phrase, VP: verb phrase, det: determiner, A: adjective, N: noun, V: verb), [8, p. 96].

According to Chomsky [9], the deep structure, the underlying relations of words with an abstract order, is not expressed but is only represented in mind, whereas surface structure is the aspect of syntactic description. Of interest here is that deep structure definition gives room for implicit meanings for words that extend the context of the subject. Therefore, although the language has some strict rules, it also has the power to enrich the thinking process with its profound structure aspects. However, words containing meanings may evoke different meanings in someone's mind when s/he extracts them. This situation affects our communication in both ways. It makes us think through various perspectives of the relevant concept, which may result in creative thinking through intended and unintended consequences. Sometimes, various meanings undermine the conversation when participants do not construct the desired meaning from the word or sentence.

Magnifying the number of decisions in the thinking process is one of the unique features of visual thinking. It brings out a different aspect of thinking: the ability to see infinitely and select the one feature designers to desire among numerous possibilities. In this sense, visual thinking is unbounded compared to language, which has syntactic and semantic rules. At every iteration, we can construct new rules and see new elements within a shape, and this process can be continuous. Figure 2 shows us that these branches can be multiplied by the rules we set every time we see them, and they do not have to follow specific rules, whereas the tree diagram of language has to. Our minds can cut a shape into different parts and create new forms or combine them to create whole new shapes while looking at it.

The flexibility and continuity of visual thinking expands the design space with many unintentional consequences of our moves. Although words can have different meanings, and meanings can change depending on who will reconstruct the idea, we are still restricted by the descriptions and their leading memories encoded in our minds. Language mediates as a conduit between how we think in our mind and how we externalize it. However, this externalization process may also provide a space for creativity since interventions made through the process cannot be defined enough. The notion of

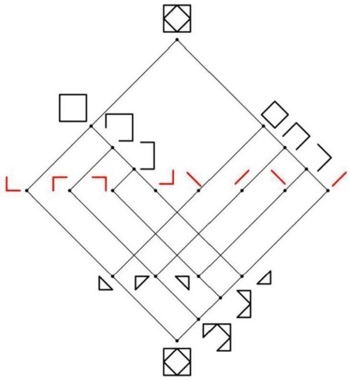


Fig. 2. A tree diagram of two squares [10].

"deep and surface structure" is common in language and visual thinking. The idea of abstract meaning only represented in mind can also be traced in visual thinking. In visual thinking, we try to externalize our ideas by employing visual elements within a constant reflective process. Here, we can emulate "visual elements" to "surface structures" and "ideas" to "deep structures." Visual thinking is also related to reflective thinking since seeing gives constant feedback to reveal the deep structure of thought.

Visual Reasoning over Digital Tools. A primary concern of cognition, particularly design cognition, is to first better understand the design to improve the design process and, secondly, to produce tools to assist designers in improving design outcomes. It is better to develop the tools for design and designers considering cognitive actions like the "seeing-moving-seeing" and contain an appreciative system comparable to a designer's appreciative system to support the evolution of the design problem [4]. The drawings are a representation of the evolving design. In that sense, the computational tools help the designers starting from the early stages to assist the evolution of the design process. Nowadays, the development of autonomous architectural tools is becoming an established research area in architecture; plan and space recognition [11]; reducing design repetitions [12]. Autonomous tools are also investigated in the field of urban design; generating urban morphologies [13]; generating building footprints and volume [14]. Bank et al. [15] also integrated GAN into their architectural design courses to explore the opportunities for designers and their role in defining the relevant domains for design possibilities, and to introduce machine learning algorithms to concept modeling processes for architectural students. Bolojan et al. [16]'s study is to apply Neural Language models (NLM) that are machine learning techniques for text-to-image synthesis by e DALL-E, VQGAN/CLIP, and Diffusion models (DMs). Text prompts with a visual reference are utilized to generate adopting John Gero's schema of 'design prototypes' [17], a new approach to architectural design was pursued, employing interconnected deep learning models. They stated that this line of work supports UN SDG #9, pushing forward the current technological capabilities of the AEC industry by offering innovative workflows.

3 A Protocol Study with AI-Powered Design Tool

The experiment consisted of two stages with the same rules and steps. Nine architects participated in the experiment individually through an online video call where each experiment was recorded as a video and audio. One out of nine architects declared that he had experienced the AI-Powered tool Midjourney (detailed explanation at Sect. 3.1) before. Then, participants (P) were provided with two architectural images (the original images): the exterior image of Ronchamp by Le Corbusier and the interior image of Therme Vals by Peter Zumthor (Fig. 3). First, the image of Ronchamp was shown to the participants, and they were asked to make verbal descriptions (T as text) of the visual in a way they felt comfortable without including the name of the building and the architect. We allow participants to use an English dictionary since English is not their native language. Each participant sent their keywords or sentences through chat to allow us to copy them to the Midjourney bot in Discord. Each participant determined their thinking duration to describe images. They were also informed that there was no limitation on how to describe the images or the number of words when asked if they should describe tangible features or the feelings the image evokes. After the first results, all participants were informed by the short presentation about how Midjourney works and how they can enrich their descriptions. Every image was shown to participants to allow them to evaluate the image and make changes accordingly. Then, each participant was given three more chances to use the Midjourney bot, including writing texts or using “Upgrade, Variations” options to reach an image similar to the first image generated. The same steps were applied for the image of Therme Vals with the chance of using the Midjourney bot three times. Each participant used the Midjourney bot seven times in total. At the end of each stage, participants were asked to decide on one of the generated images they found closer to the original image (Table 1).

Fig. 3. The original images: The exterior image of Ronchamp by Le Corbusier (left) and the interior image of Therme Vals by Peter Zumthor (right) [20].

3.1 Analyzing Tool Options

Midjourney is a text-based diffusion model; however, it offers “Variations/Upgrade” options. These are fully automated options; the user cannot control how variations/upgrade options will occur. It is important to note that choosing Variation/Upgrade

Table 1. Prompt examples from which participants made final choices

Ronchamp
P5,T1: *organic, extraordinary, looks like a shelter structure*
P6,T3: *curved black roof, curved gray roof, white house, futuristic, famous architect building, flue, small windows, grass, tree, stone building*
Therme Vals
P7, T2*: concrete, gray scale, pool, spa interior, rural landscape, tree, white illumination*
P8, T1: *perspective of a pool from inside looking at pool stairs and tiled blue-gray walls with a square openness in middle with green landscape seen from the middle, realistic style, sharp edges*
(P: Participant, T: Text)

options indicates that the user is satisfied with what s/he ends up with since these options do not generate something unexpected from the current outputs. However, our case study shows that these options do not always align with the user's mental imagery. First, we analyzed the path of choices and durations of decision-making processes for image generation from the protocol study. (Fig. 4). 'Thinking duration' for texts is determined from their typing period till their message timestamp. Thinking durations for 'Variations' and 'Upgrade' are counted as the duration from the output generation timestamp till the participants' response.

After the first results, only one participant decided to go "Variation" after writing the first text for Ronchamp, due to prior exposure to the tool. However, the participants, who are novel to the tool, needed more relevance with the first output via their first texts to the Ronchamp image. When we compare the thinking duration of T1 and T2, 50% of percent of participants spent less time during T2, and 50% percent spent more. Therefore, their reaction to their first image depends on the participant's evaluation. After the second result, except for P4, all participants wanted to write text. This time, 62% of participants increased their thinking duration. We can infer that participants realized how they describe the image can lead to the AI-powered tool many results, and not always the ones they imagine. Furthermore, after the first results, participants started to assess their words more carefully, whether the word has another meaning or how it matches with other words, or if two different words come together and stand for new meanings, or refer to different concepts. We had these inferences from the texts of Ronchamp, such as P9 wrote "*white fluke*" for text one and edited this as "*white fluke roof*" for text 2. P1 wrote "*with a blunt towe*r" for text three and edited this as "*with a blunt tower in addition to the main building*" for text four. Another example is writing "*building as a sculpture*" in addition to "*huge curved roof, curvilinear shape, thick and twisted walls*" for text three and trying to emphasize its physical attributes with the "sculpture" term. It is clear that Variations/Upgrade options were used more for the Therme Vals image. In this session, 33% of participants went for variation for the second image generation, and this ratio climbed to 77% for the third image generation, including the upgrade option. When we analyzed all tool options, no one except one participant wanted to write text after using the Variation/Upgrade option. Of interest here is that only P4, who experienced Midjourney before, went for writing after the first image variation. This supports our idea that these options do not provide a sequential image generation process.

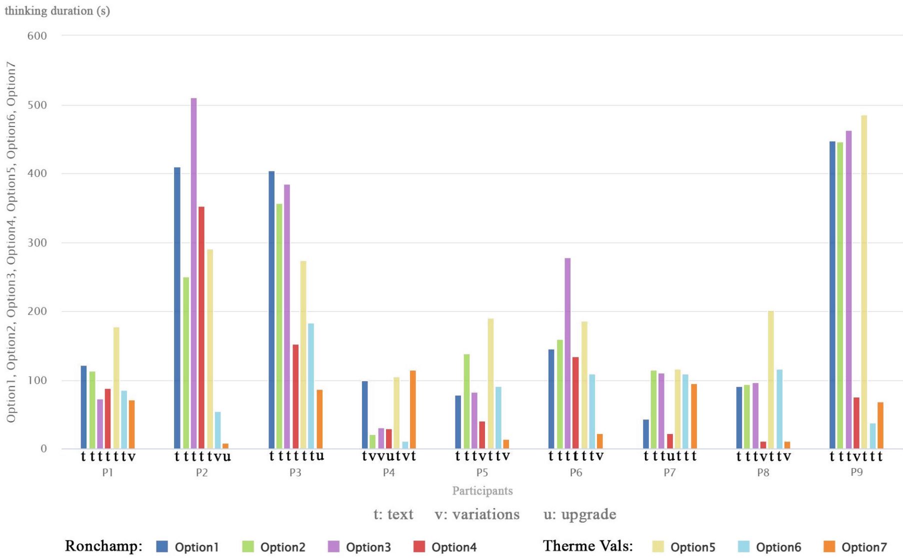
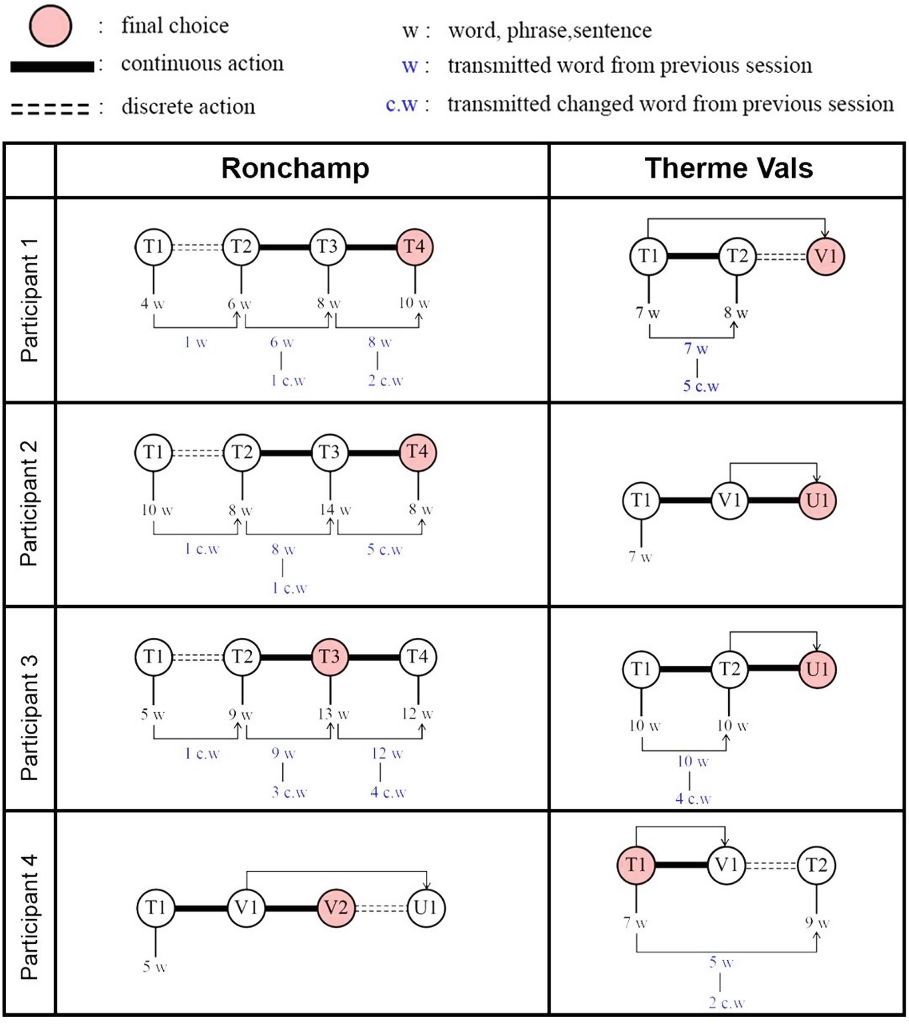


Fig. 4. Participants' option choices for description and their thinking duration(s).

3.2 Segmentation

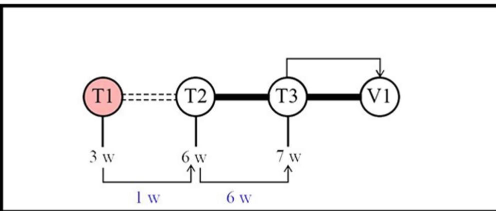
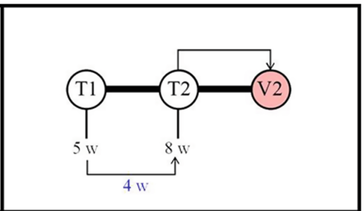
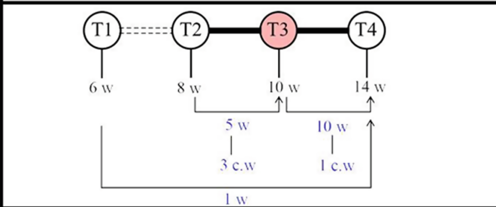
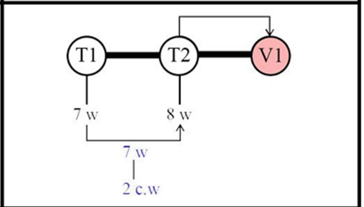
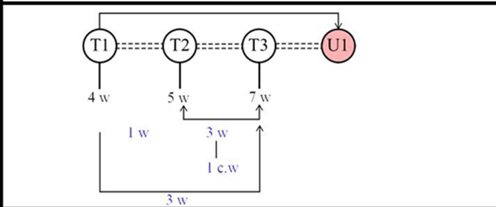
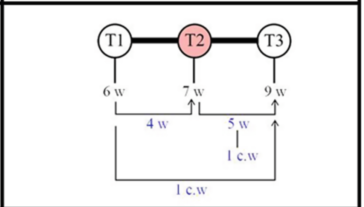
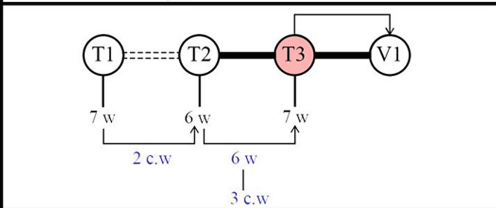
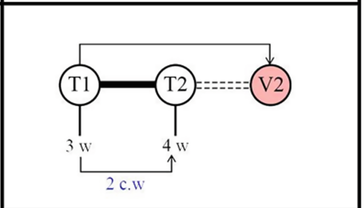
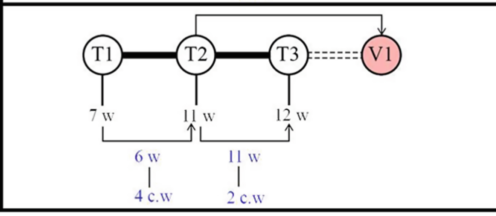
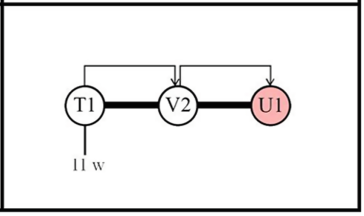
We break down the whole process into segments. For each segment, we examine participants' cognitive actions, their number of words, their transmitted words from the previous segments, and if the transmitted words were changed or not, in terms of developing the word with new words or simplifying it by reducing the words. We code the participant's cognitive action as continuous or discrete (Table 2). If a participant keeps at least 50% of the exact words from the previous text, then we code this as 'continuous' action. In this sense, participants kept their thinking process continuously, accumulating their words gradually or changing them accordingly. If a participant's half and over half of the new text consists of new words never used before, then we code this as a 'discrete' action. In this action, participants shifted their focus and tried to find new types of descriptions. Furthermore, we code Variation/Upgrade options as continuous action if these options were used based on the last image. If not, we code them as discrete actions since the user did not correspond with what s/he imagined and decided to use Variation/Upgrade options on previous actions. Finally, we highlighted the segments to which each participant's final choice belongs so we could make inferences about their experience. When we analyzed their actions based on our coding scheme, 77% of participants made discrete actions when they wrote T2 for Ronchamp. We can infer that they developed new ideas after they evaluated their results and got information about the tool Midjourney. For instance, when we analyzed T1 and T2 for Therme Vals, none of the participants made discrete actions. In fact, only 3 participants made discrete actions during the Therme Vals session, and these actions were made for the last option. In this sense, it is clear that the more participants are exposed to the tool, the longer they would sequentially keep their thinking. 66% increased their number of words for T2 of Ronchamp, and 83% also made discrete actions. Therefore, we can conjecture that participants thought they would be more likely to get similar results if they elaborate their descriptions with more words. When we asked participants to choose one result that

Table 2. Segmentation of participants' design actions. (T: text, V: variation, U: upgrade).



(*continued*)

Table 2. (*continued*)

Participant 5		
Participant 6		
Participant 7		
Participant 8		
Participant 9		

they think is the most similar to the original image, the choices were interesting. They did not always choose what they ended up with or the image they produced with more words. P5 chose one of the images he produced with T1 for Ronchamp, where he had no idea what Midjourney was. 44% of participants did not choose the last image they generated for Ronchamp. For Therme Vals, 77% chose the last image they generated. We can infer that as designers engage more with a new tool, such as Midjourney, it changes the decision processes of the designers.

We measured perceptual distance, which measures how similar two images are in a way that coincides with human judgment [18]. We used the code available on Github [19] that the author provided and we executed it on Google Colab. The results of metrics range between 0 to 1 with being 0 means they are the same (Fig. 5). The lower the number, the greater the similarity. P4 has the most similar result with 0.598 for Ronchamp and P2 has it with 0.637 for Therme Vals (Fig. 6). The results are not in correlation with educational background since P6's final choice, who is a doctorate student and a practicing architect, is rated as 6th the most similar image for Ronchamp whereas P4's, who is a master student with no practice experience, is rated as 1st one. P2's final choice, who is a practicing architect with 2 years of experience and without a graduate degree, is rated as the most similar one for Therme Vals while P9's choice, who is both an architect and software engineer with two major degrees, is rated as 7th among others.

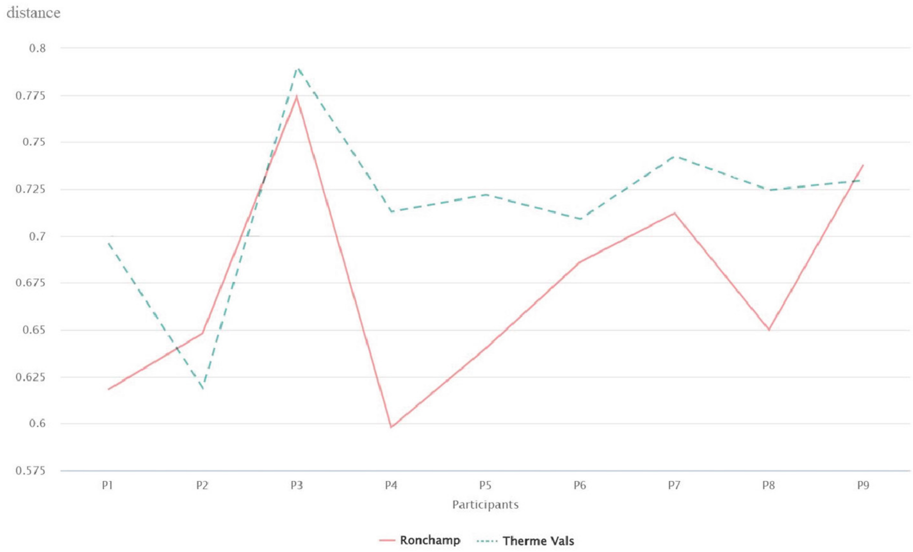


Fig. 5. Perceptual Metrics of final selected images.

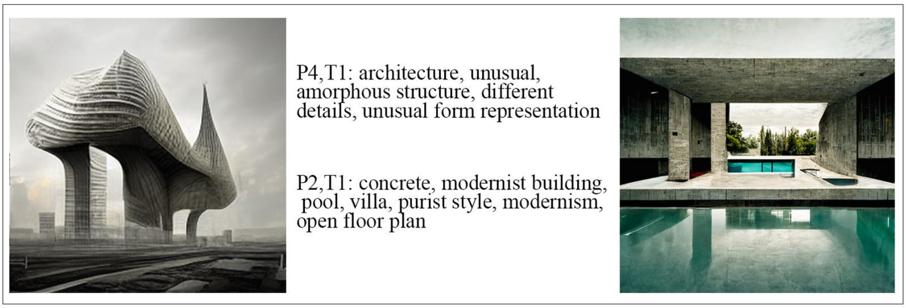


Fig. 6. The most similar images are based on the perceptual similarity metric.

3.3 Encoding Descriptions into Categories

Participants delivered different descriptions that emphasized images' various aspects during the experiment. In order to analyze how descriptions differ, we devised seven categories to encode all the information that participants provide through text writing. These categories include type, physical attributes, style, material, feeling, reference, and render style. The six categories were shown to participants except for physical attributes to enrich their descriptions during the presentation. The "Physical attributes" category was devised after the experiment when we examined descriptions and realized many physical descriptions exist. The encoding we developed here explores architects' architectural description characteristics by analyzing what categories the majority of words belong to. Since the number of discrete actions was the highest during writing text two for Ronchamp, we wanted to demonstrate how participants' word choice changed after they evaluated their results generated with T1. To clarify how we encode descriptions, we showed some examples and the output image of texts (Table 3). Figure 7 shows the categories and how frequently they are preferred by participants for T1-T2 of Ronchamp and Therme Vals. If no words exist for each category, the user selects the Variation/Upgrade option. It is quite noticeable that participants mostly use words related to physical attributes both for Ronchamp and Therme Vals. For T1-T2 of Therme Vals, physical attributes descriptions are dominant, whereas T2 only consists of physical attributes. Furthermore, 88% expressed their feelings for text one of Ronchamp, while this ratio declined by 33% for T2 even though we informed participants they could include their feelings.

4 Discussion and Conclusion

The case study shows that participants' first impressions of the image and how they express their impressions through descriptions do not correspond with how an AI-powered tool (in our case, Midjourney) interprets those descriptions. Furthermore, Variations/Upgrade options do not always result in compatible images even though they are programmed to be as. This incompatibility may arise for two reasons: AI-powered tools are in their nascent phases to assist architects or architects need to adapt to communicate with an AI-powered tool properly. The results show that generated images do not always get better based on the number of words or how developed the words are, and participants tend to describe the images' physical attributes more than the other categories. However, in the study we conducted using Midjourney shows us that describing only physical attributes is not enough for desired results. Abstract concepts such as feelings, style and references help both the AI-powered tool and the participant to elaborate their textual definitions.

Digital external representation tools usually have been developed based on the insights of sketches. AI-powered tools such as Midjourney bring a new design space where users externalize their mental imagery with verbal descriptions. When architects were restricted to draw, they struggled to control the design process which they would amplify with "see-move-see". Instead of adopting a heuristic approach, they experienced a stochastic approach as recalling the most common and appropriate words to prevent AI from being puzzled. We see that architects still adopted a reflective thinking process

Table 3. Encoding descriptions into categories

P2: text 2-Ronchamp: chapel, France, white walls, purism, curvilinear shape, dramatic effect, modernist architecture, curved roof type: chapel physical attributes: white walls, curvilinear shape, curved roof style: modernist architecture feeling: dramatic effect,purism reference: France	
P8: text 1-Ronchamp: organic, spacious, sharp, bright, brave, geometric, plain physical attributes: organic, sharp, bright, geometric, plain feeling: spacious, brave	
P9: text 1- Therme Vals: modern half-open space, pool in the half-open space, geometrical sharpness, geometrical balance, modernist, modernist simplicity, brutal concrete, relaxing, forest view, heidegger architecture philosophy, interior half-open wet space type: modern half-open space, interior half-open wet space physical attributes: pool in the half-open space, geometrical sharpness, geometrical balance, forest view style: modernist material: brutal concrete feeling: relaxing reference: modernist simplicity, heidegger architecture philosophy	

with the cycle where they adjust their descriptions based on outputs. How Midjourney interprets the word can also give rise to creative thinking since the deep structure in language depends on the one who extracts the meaning. However, using only words and relinquishing the drawing seems not quite robust enough to externalize the mental imagery of architects. AI-powered tools should be developed to corroborate the shared mode of interaction with architects. Furthermore, the drawing process is more internal in that architects evaluate most of the ideas implicitly in their minds. However, AI-powered tools want us to externalize our thinking process explicitly through language. We can imply that exploiting this kind of AI-powered tool at its best depends on how much tacit knowledge we can transform into explicit knowledge, which can be expressed through language.

The change in the descriptions during the experiment gave us a hint of how the designers can describe architectural sceneries and how they try to understand the perspective of the AI-powered tool as it has its knowledge. This change in the mental mode is one of the most important points of the reflection process due to its collaborative nature with an AI. Even though the tool used in this experiment is not unbiased, its collaboration with the designers is unique at every sequence due to the interaction both parties create during the design process.

In the protocol study, the interaction was not direct, and the designers were getting familiar with the process. However, due to the tool's novelty, they needed some time to grasp the dynamics of the process. Despite the lack of direct interaction with the tool, the

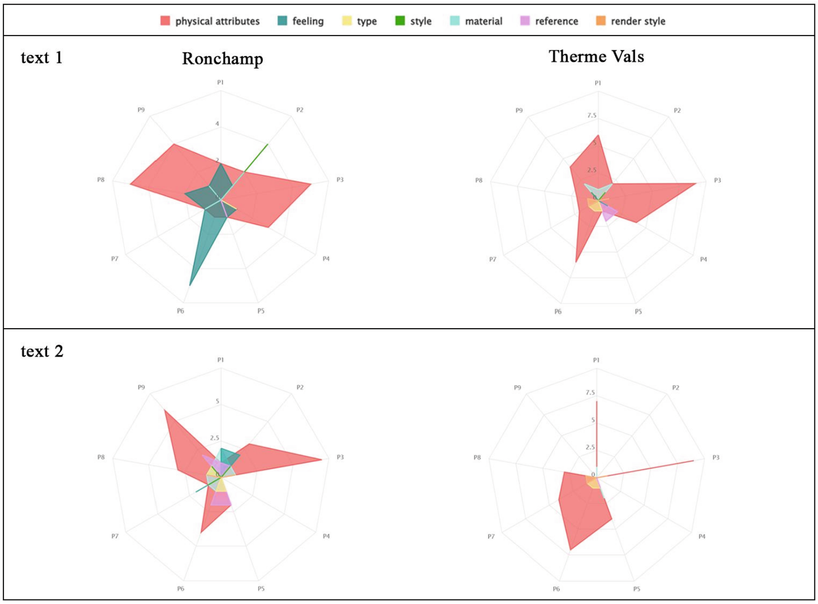


Fig. 7. The categories and how frequently participants prefer them.

design process allowed us to evaluate the wording designers use during the generative processes. One major finding was that the design process's nonlinearity was still present. Even though the AI-powered tool generated endless variations, the decision-making process was the human designer. The final decisions differed from the latest images generated and were the most similar to the original images that participants think as they were. This also shows the divergence between computational and design thinking; ambiguity and aesthetics play a different role for the human designers than the AI model, like GAN and stable diffusion. Another finding is that the wording is another skill for designers, and how they define scenery is unique to their prior architectural education, literature, and mother language. Midjourney is developed for the English language, and the study was conducted with non-native English speakers; this may have limited the participants' expression of feelings and other descriptive words even though they were allowed to use a dictionary.

The developing tools with AI models are changing how we interact with digital tools, hence the way we design using these tools. The increase of voice-commanded devices we use in our daily lives, such as smartphones and home appliances, changes how we interact with our environment. Therefore the interaction with smart design tools will affect how we interact with the new design tools. The changing modes of interactions also have the potential to transform the roles of tools in design processes. The knowledge and the support that AI-powered tools will allow designers to recognize them as collaborative entities that help designers to decide with an extended intelligence.

Acknowledgements. We would like to thank Prof. Mine Özkar for her insightful course "Computing Theories and Models in Architectural Design" given at ITU. "Reasoning through Visual and Language" section has been developed through this course.

References

1. Hay, L., Cash, P., McKilligan, S.: The future of design cognition analysis. Des. Sci. **6**, e20 (2020). https://doi.org/10.1017/dsj.2020.20
2. Goldschmidt, G.: The dialectics of sketching. Creat. Res. J. **4**(2), 123–143 (1991). https://doi.org/10.1080/10400419109534381
3. Arnheim, R.: Sketching and the psychology of design. Des. Issues **9**(2), 15 (1993). https://doi.org/10.2307/1511669
4. Schon, D.A.: Designing as reflective conversation with the materials of a design situation. Res. Eng. Design **3**(3), 131–147 (1992). https://doi.org/10.1007/bf01580516
5. Arnheim, R.: A plea for visual thinking. Critical Inquiry **6**(3), 489–497 (1980). University of California Press
6. Goldschmidt, G.: On visual design thinking: the vis kids of architecture. Des. Stud. **15**(2), 158–174 (1994). https://doi.org/10.1016/0142-694x(94)90022-1
7. Bilda, Z., Gero, J.S.: Does sketching off-load visuo-spatial working memory. Studying Designers **5**, 145–160 (2005)
8. Pinker, S.: The Language Instinct, p. 96. W. Morrow and Co., Harper Collins Publishers, New York (1994)https://doi.org/10.1057/9780230379411_23
9. Chomsky, N.: Cartesian Linguistics. Harper, and Row, New York, London (1966)
10. Stiny, G.: The critic as artist: Oscar Wilde's prolegomena to shape grammars. Nexus Netw. J. **17**(3), 723–758 (2016)
11. Huang, W., Zheng, H.: Architectural drawings recognition and generation through machine learning. In: Proceedings of the 38th ACADIA, pp. 18–20. Mexico City, Mexico (2018)
12. Uzun, C., Çolakoğlu, M.B., İnceoğlu, A.: GAN as a generative architectural plan layout tool: a case study for training DCGAN with palladian plans and evaluation of DCGAN outputs. A| Z ITU J. Faculty Architect. **17**, 185–198 (2020)
13. Boim, A., Dortheimer, J., Sprecher, A.: A machine-learning approach to urban design interventions in non-planned settlements. In: van Ameijde, J., Gardner, N., Hoon Hyun, K., Luo, D., Sheth, U. (eds.) POST-CARBON - Proceedings of the 27th CAADRIA Conference, pp. 223–232 (2022)
14. Di Carlo, R., Mittal, D., Vesely, O.: Generating 3D building volumes for a given urban context using Pix2Pix GAN. In: Proceedings of the 40th Conference of eCAADe, pp. 287–296 (2023)
15. Bank, M., Sandor, V., Schinegger, K., Rutzinger, S.: Learning spatiality-A GAN method for designing architectural models through labeled sections. In: Pak, B., Wurzer, G., Stouffs, R. (eds.) Co-creating the Future: Inclusion in and through Design - Proceedings of the 40th Conference of eCAADe, vol. 2, Ghent, 13–16 September 2022, pp. 611–619 (2022)
16. Bolojan, D., Vermisso, E., Yousif, S.: Is language all we need? A query into architectural semantics using a multimodal generative workflow. In: Proceedings of the 27th CAADRIA Proceedings (2022). https://doi.org/10.52842/conf.caadria.2022.1.353
17. Gero, J.S.: Design prototypes: a knowledge representation schema for design. AI Mag. **11**(4), 26 (1990)
18. Zhang, R., Isola, P., Efros, A.A., Shectman, E., Wang, O.: The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. arXiv (2018). https://doi.org/10.48550/arXiv.1801.03924
19. Github. https://github.com/richzhang/PerceptualSimilarity. last accessed 2023/02/21
20. Archdaily. https://www.archdaily.com/84988/ad-classics-ronchamp-le-corbusier. Vals. https://vals.ch/erleben/erholung/therme-vals/. last accessed 2023/02/28