USING TEXT-TO-IMAGE GENERATION FOR ARCHITECTURAL DESIGN IDEATION

A PREPRINT

Ville Paananen
Center for Ubiquitous Computing
University of Oulu
Oulu, Finland
ville.paananen@oulu.fi

Jonas Oppenlaender
Faculty of Information Technology
University of Jyväskylä
Jyväskylä, Finland
jonas.x1.oppenlander@jyu.fi

Aku Visuri
Center for Ubiquitous Computing
University of Oulu
Oulu, Finland
aku.visuri@oulu.fi

April 20, 2023

ABSTRACT

The recent progress of text-to-image generation has been recognized in architectural design. Our study is the first to investigate the potential of text-to-image generators in supporting creativity during the early stages of the architectural design process. We conducted a laboratory study with 17 architecture students, who developed a concept for a culture center using three popular text-to-image generators: Midjourney, Stable Diffusion, and DALL-E. Through standardized questionnaires and group interviews, we found that image generation could be a meaningful part of the design process when design constraints are carefully considered. Generative tools support serendipitous discovery of ideas and an imaginative mindset, enriching the design process. We identified several challenges of image generators and provided considerations for software development and educators to support creativity and emphasize designers' imaginative mindset. By understanding the limitations and potential of text-to-image generators, architects and designers can leverage this technology in their design process and education, facilitating innovation and effective communication of concepts.

Keywords Architecture, text-to-image generation, generative AI, design creativity

1 Introduction

In the field of architecture, effective ideation hinges on the ability to represent ideas. Traditionally, drawings, photographs, and other visual media have been used to stimulate ideation and communicate design concepts. However, recent advances in generative artificial intelligence (AI) have made it possible to generate detailed and realistic representations of architectural concepts, using prompting in natural language as a general-purpose interface [1, 2, 3, 4, 5]. The use of generative AI and procedural design is not new to architecture, and their use date back to the 1970s [6]. However, prompt-based generation marks a paradigm shift that could affect the architectural design process. Text-to-image generation tools [3, 7] are one such example of generative AI. Text-to-image generation tools can allow for a quick conceptualization of ideas with natural language during the idea generation process. Thus, these tools have the potential to transform the way architects and designers develop and communicate their ideas.

In this paper, we study how different text-to-image generators can support creativity during the "fuzzy front end" [8] of new concept development in the early stages of the architectural design process. In particular, we investigate:

1. How can text-to-image generators support creativity and ideation during the early stages of architectural design?
2. How effective are out-of-the-box text-to-image generators in the context of architectural design, and what future considerations could developers take into account?
3. What are the typical challenges of text-to-image generator use and text prompting for novel users?

In a laboratory study, 17 participants developed a concept for a culture center using three popular text-to-image generators; Midjourney [9], Stable Diffusion [3], and DALL-E [7]. Through standardized questionnaires on creativity support tools and group interviews, we learned that image generation could be a meaningful part of the design process when design constraints and imaginative ideation are carefully considered. Generative tools support the serendipitous discovery of ideas and an imaginative mindset, which can enrich the design process. Through our study, we highlight several challenges of image generators and provide considerations for software development and educators to support creativity and emphasize designers' imaginative mindset.

2 Related work

To contextualize the use of text-to-image generators in architectural design, we first describe the different ways creativity has been approached in the architectural design process. Then, we present current literature on how creativity-supporting generative tools have been applied in the architectural design process.

2.1 Creativity in the architectural design process

Architecture is highly relevant to creativity as a field concerned with solving problems in contextual and effective ways [10] The prior literature has developed multiple notions for understanding what aspects affect creativity and how it can be supported in architectural design. For instance, Sarkar and Chakrabarti formulated creativity as the function of novelty and usefulness, allowing for the assessment of different design outcomes [11]. This two-factor definition of creativity into novelty (or its related synonyms, such as originality, unusualness, or uniqueness) and usefulness (or its related synonyms effectiveness, fit, or appropriateness) – has become popular in scholarly literature as a way to operationalize and measure the concept of creativity [12]. However, creativity in architecture is not only related to producing a final outcome that is novel and useful but it is also related to the application of one's creative skills during the creative design process. Casakin et al. found that creative thinking skills are more related to verbal skills rather than figural skills [13]. Consequently, the authors proposed that the creative skills in architecture are also generalizable to other problem-solving areas in life. Additionally, Baghai Daemei & Safari found that students' experience of design processes is a critical aspect of creativity [14].

Creativity has also been operationalized in architecture using various tools. Park et al. studied how text stimuli can improve a person's imagination in nonlinear architectural design tasks [15]. Using text snippets from Italo Calvino's *Imaginary Cities*, the partaking architecture students produced imaginative concepts. The findings show that text stimuli can support nonlinear creativity and move the design focus from the outcome to the process. As such, the multimedia approach requires designers' imagination. In order to assess artifact creativity, Demirkan & Afacan used exploratory and confirmatory factor analysis to develop a descriptive set of 41 design creativity words [16]. The design elements were then grouped into three factors: "Artifact creativity," "Design elements," and "Assembly of creativity elements." Kowaltowski et al. interviewed architecture educators about stimulating creativity in architectural design [17]. The resulting creativity support methods show that generative methods are relevant for producing many unfamiliar ideas, but it also requires more support from teachers. Similarly, Mose Biskjaer et al. developed an analytical framework for understanding how creativity methods are used in design processes [18]. The authors suggest that the design process itself can be designed through concrete, conceptual, and design spaces aspects. As such, reformulating the design task can help to bring forward more creative solutions.

2.2 Text-to-image generation in architectural design

Text-guided diffusion models [1, 2, 3, 4] have become a popular means of synthesizing novel images from input prompts written in natural language, and generative AI is increasingly being employed in academia and in the industry. Generative AI in architectural design has been explored in two surveys. Through reviewing machine learning research trends in architecture, Ozerol & Arslan Selçuk found that generative AI is rising in popularity [19]. However, machine learning was more often applied in 3D generative methods than in 2D. This suggests that the fast development of text-to-image generative methods has not yet reached the architectural research community. In a survey of generative systems in architectural, engineering, and construction research between 2009 and 2019, BuHamdan et al. found that many generative methods are focused on architectural, structural engineering, and urban design disciplines [20]. However, the most popular architectural use cases (facade design, form generation, layout generation) represent more geometric processes, and the role of more conceptual creativity in generative systems is still left to be unexplored.

Text-to-image generation systems provide an easy-to-use interface due to the ability to respond to natural language prompts. However, the creativity of text-to-image generation currently still hinges on the skill of its users [21]. To control the output, users have to resort to special keywords in the prompts to produce images in a certain style or

quality [22]. Longer prompts also typically produce images of higher quality [23]. While text-to-image generation tools can be intuitive, their application in the context of architecture remains yet to be explored.

Seneviratne et al. used a systematic grammar to explore the robustness of a text-to-image generator in the context of the built environment [24]. The study found that the image generator was broadly applicable in the context of architecture. However, architectural semantics contain ambiguities [25], and the real-world benefit of text-guided image generation remains to be explored. In this paper, we specifically focus on how text-to-image generation tools can support human divergent creativity during the early-stage concept design process.

3 Method

We designed a laboratory study in which architecture students engaged with text-to-image generators in a short architectural design task. The study was tested in a formative pilot study in which two authors and three colleagues used text-to-image generators to create their dream home. The formative pilot informed the design of the main study design, as follows.

3.1 Study design and procedure

We conducted three sessions (henceforth **S1**, **S2**, and **S3**) where 5–6 participants each individually worked on the same task. Participants were tasked to design a concept for a culture center. The site is a small island with a small observatory and interconnecting bridges to the nearby downtown area. Participants were tasked to brainstorm visual concepts supporting the overall design task. More specifically, participants were asked to produce visual representations of their concept, including 1) a floorplan, 2) an interior perspective visualization, and 3) a facade material sample. Participants could use pen and paper to support their ideation, but the generated images could not be edited digitally.

The study procedure was as follows. Participants were first asked to provide informed consent. Participants were then given an introduction to text-to-image generators. The text-to-image user interfaces were not modified in any way. Participants were introduced to the tools with a short presentation on how to use the text-to-image generators for architecture and some unrelated example images produced by each of the three tools. Participants were then given a short interactive tutorial task that revolved around generating and iterating an increasingly complicated image of a pineapple. Only basic functionalities – text prompts and generating variations or upscaled versions of generated images – of the tools were allowed during the session to ensure comparability of the three tools. Advanced features that participants were not allowed to use during the study include features like inpainting or using generated images as the basis for future generations. These features were omitted since not all of the three tools contain all advanced features. Participants were then presented with a design brief and began working on the task.

Participants had between 1h 15min and 1h 25min to work on their task. The duration varied according to how long the initial stages of the study lasted and at what point the participants felt they had completed the task. In each session, all three tools were used, with 1–2 participants each using one of the three tools. Each participant had their own laptop (either personal or borrowed) to work with. While they worked individually towards their own solution, participants could talk and discuss freely amongst each other, as well as take breaks when needed. This decision was taken to emulate a collaborative work environment in an organization and to support the participants' individual creative needs. One researcher was present at all times to answer questions. A second researcher acted as an observer, taking notes of any interesting discussions and observations during the design session. The sessions were recorded using a conference microphone that could capture the audio in the whole meeting room. Once participants finished working on the design task, each participant presented their work to the group, and to further motivate the participants, the participants voted for the best work using ranked choice and ranking their own design lowest. Participants were compensated with a 15 EUR gift card and the winner of each session was awarded an additional 15 EUR gift card.

3.2 Data collection

At the end of the session, we administered the Creativity Support Index (CSI) [26] to evaluate how the image generators supported the participants' creative processes. Based on the recommendations by Cherry and Latulipe [26], collaboration was marked as an optional item. This approach was adopted as participants were allowed to collaborate, but some opted not to do so actively. After the ideation sessions, we conducted semi-structured group interviews focused on three main aspects: 1) how well the tools could produce the required images, 2) whether the tools provided novel solutions and 3) how participants thought the tools could be used in their design practice. As we learned more about the tools in the first session, we also focused on participants' comments about an "ideal" tool that would support their design tasks. Two researchers conducted the interview, with one leading the discussion and another acting as a scribe, taking comprehensive notes. The participant's comments during each session were written down by a researcher during the

Table 1: The list of 17 study participants and their ages, genders, self reported years of studying architecture, study session, image generation tool, and prior experience with image generation.

No.	Age	Gender	Years studied	Session	Tool	Experience
P1	22	Male	Three	S1	DALL-E	Yes
P2	21	Female	Three	S1	DALL-E	No
P3	21	Female	Three	S1	Stable Diffusion	Yes
P4	29	Female	Two	S1	Stable Diffusion	No
P5	21	Female	Three	S1	Midjourney	No
P6	26	Female	Five	S2	DALL-E	No
P7	47	Male	More than five	S2	DALL-E	Yes
P8	24	Female	Three	S2	Stable Diffusion	Yes
P9	29	Female	Seven	S2	Stable Diffusion	No
P10	24	Male	Four	S2	Midjourney	Yes
P11	43	Male	One	S2	Midjourney	Yes
P12	25	Female	Five	S3	DALL-E	No
P13	25	Male	Five	S3	DALL-E	Yes
P14	23	Female	Four	S3	Stable Diffusion	No
P15	24	Male	Five	S3	Stable Diffusion	Yes
P16	23	Female	One	S3	Midjourney	No
P17	24	Female	Three	S3	Midjourney	No
	M = 27	35.3% Male				52% No
	StDev = 7	64.7% Female				48% Yes

interview, and complemented with transcribed audio recordings. The full commentary consists of 2905 words and the audio recordings are approximately 50 minutes in total. The commentary was analysed using content analysis [27] to identify the participants' experiences using the tools.

3.3 Image generation tools

The study was conducted with three text-to-image generation tools: Midjourney (version 4; **MJ**), DALL-E 2 (**DE**), and Stable Diffusion (version 1.5; **SD**). These three tools represent the current state-of-the-art of text-to-image generation accessible to the public. The tools have become very popular as they provide an easy-to-use means of synthesizing images from written text prompts. Midjourney and DALL-E are provided as web-based services by Midjourney and OpenAI, respectively.[1] For Stable Diffusion, we used the web-based *Dream Studio* interface,[2] as provided by Stability AI. The use of each of the three systems (MJ, DE, SD) in the sessions was balanced between participants, with up to two participants individually using one of the tools in each session. Each account at the three image generation services was filled with sufficient credits to allow image generation during the session. Participants were allowed to use their own laptops, apart from participants assigned to use SD. Data from two SD participants (P3, P4) was lost during S1 when it was noticed that SD only stores prompt history in the participant's local browser history. Participants assigned to use SD in S2-S3 worked with a provided laptop that allows us to access the locally stored browser history. The data from S2-S3 is limited to only 100 of the last prompts as the SD does not store a complete history of prompts.

3.4 Participants

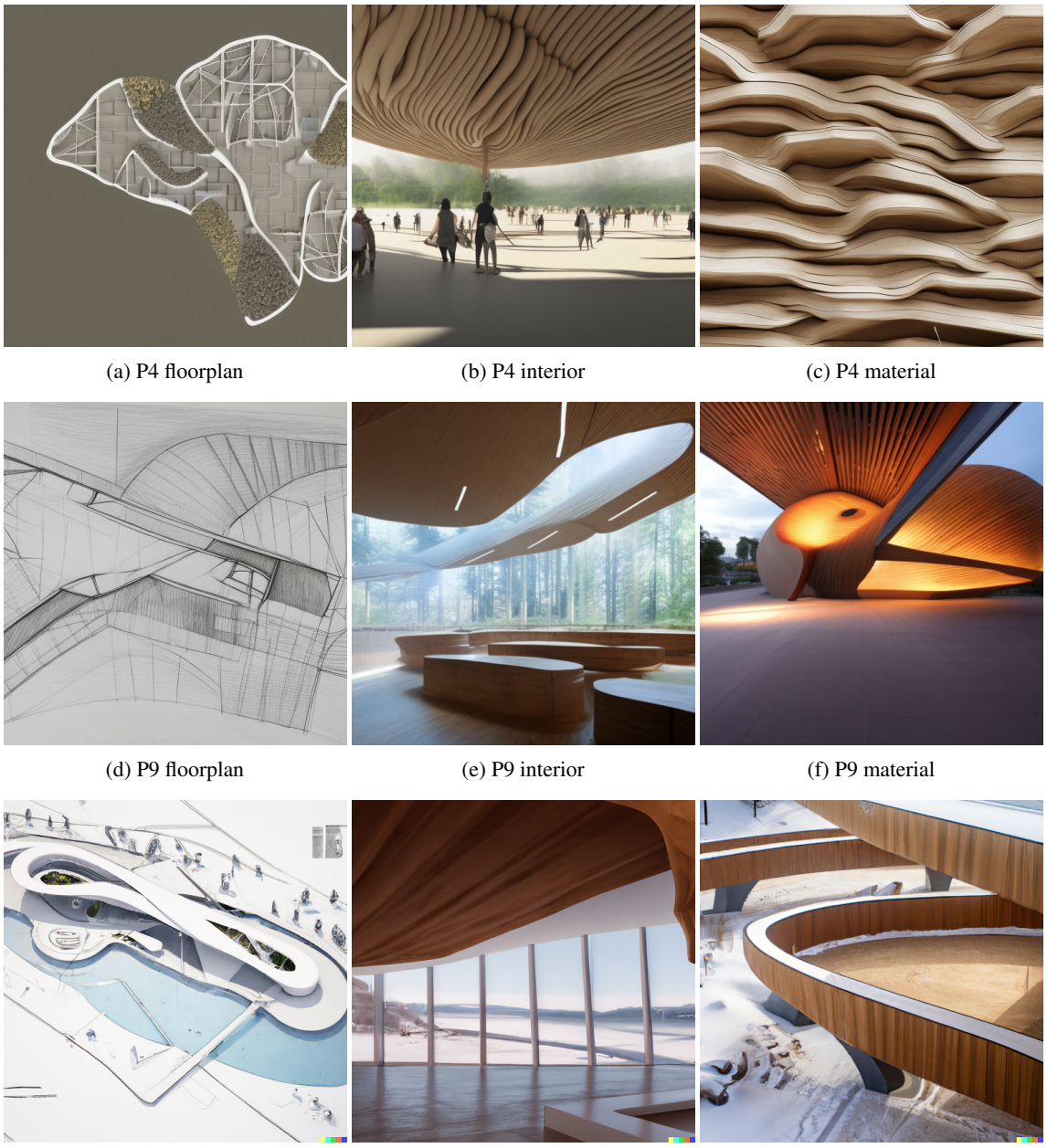
We recruited 17 participants (P1–P17; 11 women, 6 men; self-reported genders) of ages 21–47 ($M = 27$ years, $StDev = 7$ years) using a university mailing list and instant messaging channels for architecture students. All participants were first to seventh-year architecture students, with a majority (35%) being third-year students. Nine out of the 17 participants (52%) had not used image generators previously, and only three (18%) participants reported having used the systems five or more times. Seven participants with prior experience stated they had used the tools just for fun and testing. Only one person said they had used it for visualization tasks. See Table 1 for the full description of the participants.

[1] https://www.midjourney.com, https://labs.openai.com

[2] https://beta.dreamstudio.ai

4 Results

In the three sessions, participants produced images with diverse prompts. In the following, we describe the generated images, analyse the prompt language used by the participants, and then the interview data, and general feedback during the sessions. In the qualitative section, we evaluate the efficacy of the image generators in facilitating the design task, examine the participants' utilization of prompts to visualize their ideas, and discuss the qualitative insights gleaned from the group interviews.



(a) P4 floorplan (b) P4 interior (c) P4 material

(d) P9 floorplan (e) P9 interior (f) P9 material

(g) P12 floorplan (h) P12 interior (i) P12 material

Figure 1: The three participants' floorplans, interior views, and facade materials voted best works from their respective sessions S1–S3.