

THE ROLE OF ARTIFICIAL INTELLIGENCE IN ARCHITECTURAL VISUALIZATION

DOI: https://doi.org/10.18485/arh_pt.2025.10.ch61

_ Nevena Filipović

PhD student, Teaching Assistant, Faculty of
Architecture - University of Belgrade, Belgrade, Serbia,
nevena97filipovic@gmail.com, [0009-0000-8024-8046]

_ Vladimir Kovač

PhD, Assistant Professor, Faculty of
Architecture - University of Belgrade, Belgrade,
Serbia, kovac@arh.bg.ac.rs, [0000-0003-2117-8740]

_ Danira Sovilj

PhD, Teaching Assistant, Faculty of
Architecture - University of Belgrade, Belgrade,
Serbia, sovilj@arh.bg.ac.rs, [0000-0002-4294-9244]

ABSTRACT

The paper explores and discusses the contemporary position of architectural visualization in the context of the expansion of artificial intelligence. The question arises whether artificial intelligence represents a fundamental disruption to an architect's design process or is a powerful, time-saving tool. The research involves a systematic literature review of text-to-image-based software and research by design using image-to-image-based AI software.

Here, it is shown to what extent different phases of architectural presentation can be replaced with AI software. This paper primarily focuses on a stage of the creative process where abstract geometric form is transformed into architecture. The paper explores the possibilities of AI in bridging the gap between abstract geometry and architecture as realistic, lived space. We claim that this ability is one of the core parts of being an architect. Thus, we examine whether AI can replace an architect in this matter. We suggest that image-to-image-based software shows better spatial reasoning than text-to-image models due to integrated structure match functions. However, they cannot wholly replace the architect.

The findings reveal the importance of an architect's curatorial role in creating architectural drawings using AI software. Furthermore, the limits and potential in spatial reasoning of text-to-image and image-to-image-based AI tools are being defined. Bearing in mind that architectural drawing has always absorbed all the technological patterns of the time in which it was created, the immediate goal of the research is to define those aspects that make it a universal tool of architecture in the era of artificial intelligence.

KEYWORDS _ *Architectural visualization, architectural presentation, architectural drawing, text-to-image model, image-to-image model, artificial Intelligence in architecture*

INTRODUCTION

The application of artificial intelligence in architecture is based on computer programs that mimic human cognitive processes at various stages of the design process. When we consider architectural visualization part of an architect's design process, we recognize the two most common phases in which artificial intelligence is used. The first is a visualization of the initial idea, an abstract concept that explains the project's atmosphere without defining spatial relationships. Research has shown that text-to-image models have high potential in this initial design phase, creating various solutions and providing inspiration [6] [8] [14]. The second is the final phase of a project, where artificial intelligence is used to visualize the final solution, bridging the gap between sketches and photorealistic renderings. [7] Recently, an increasing number of software using image-to-image models to transform sketches or linear drawings into realistic renders has become available online [1] [4] [12], claiming to become the predominant way of creating renders.

We recognize that the use of artificial intelligence for architectural visualization is focused on the beginning, when there is only a blurry idea of the project, and the end of a design process, where almost every detail is defined. However, most of the creative process is happening in between. The path from idea to realization is considered an architect's mastery. This paper focuses on a specific stage of the design process where abstract geometric form is transformed into architecture. The paper explores the possibilities of artificial intelligence in bridging the gap between abstract geometry and architecture as realistic, lived space.

We consider abstract geometric form as a group of objects (volumes) that are in specific relation to one another. For example: "Three white cubes stacked on top of each other. The bottom is the largest, and the middle is the smallest." This describes a geometric composition that could be used as a spatial concept for a house. In human perception, the fundamental relations (e.g., in, on, under, near) are perceived as directly as basic object properties [5], making relations basic building blocks of human cognition [3]. However, research shows that AI software performs poorly in spatial relationship understanding skills [2] [3] [9] and thus proves insufficient in human-level visual reasoning [2] [8]. Whereas the visual quality of the images can be stunning [8] [9], their recognition of form is much worse [8] [15].

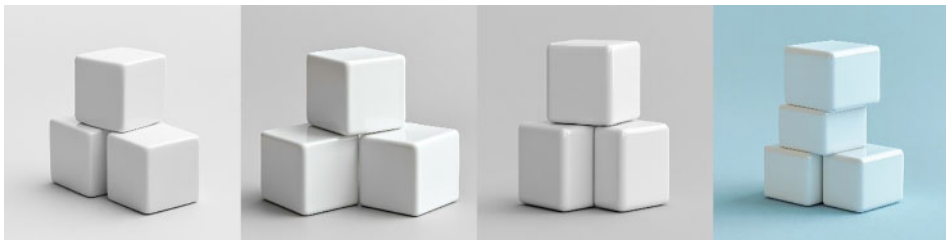


Figure 1: Images generated with Midjourney using the prompt: "Three white cubes stacked on top of each other. The bottom one is the largest and the middle one is the smallest."

We claim that the ability to visualize abstract geometric forms as lived space is one of the core parts of being an architect. Thus, we examine whether AI can replace an architect in this matter. Previous research shows that text-to-image AI software performs poorly in spatial reasoning [2] [3] [8] [9]. Thus, we extend the research to the image-to-image model. We are not testing the understanding of fundamental spatial relations but rather basic architectural logic – the relations between elementary architectural elements (materials, lighting, windows, doors, stairs) and space. We suggest that image-to-image models show better spatial reasoning than text-to-image models due to the integrated structure match function. However, they cannot wholly replace the architect. We indicate the need for a co-creative process between human and artificial intelligence [13], where the architect's curatorial role in selecting material and directing the process is crucial.

IMAGE GENERATION MODELS

Text-to-image models were initially based on large datasets of annotated images. The annotation process works by offering annotators images and text descriptions, nouns, and phrases they need to connect. This was the subject of criticism since sometimes we need more words, including verbs and adjectives, to explain something we see. As in real life, when we see a photograph, we often understand it as a dynamic representation of events and relations between objects, not so much as a static sum of objects. We explain what is happening in the photograph, as opposed to annotations that speak exclusively about what is in the photograph. Although text-to-image models are rapidly improving with expanding datasets and new benchmarks, there are still shortcomings in considering (spatial) reasoning. The difficulties regarding text prompts [6] [7] [8] [11] prompted the development of AI software that combines image and text inputs for generating new visuals. These transitional models led to the creation of an image-to-image model, primarily image-based.

Image-to-image models generate a new image based on an input image. These models can be supplemented with text prompts directing the software to the crucial aspects of the input image. Much such software has integrated additional settings - structure recognition, multiple input images, face recognition, and integrated styles (e.g., hand drawing, realistic photography, illustration). All integrated settings further direct how the software generates a new image and can be decisive in achieving the desired results.

RESEARCH BY DESIGN

Previous research has shown the possibility of applying the text-to-image model for creating concepts, inspiration, and variations of solutions in the initial design phase [6] [8] [14]. However, it has difficulties recognizing spatial relations and form [2] [3] [9] [14] [15]. We found that some text-to-image models can imagine architecture from a text description of an abstract geometry but cannot interpret that exact geometry as a building.



Figure 2: Images generated with DALL-E through ChatGPT using the prompt: “Three white cubes stacked on top of each other. The bottom one is the largest and the middle one is the smallest.” for the left image. The software was then asked “Can you imagine this as architecture?” and the right image was created.

Thus, this research uses an image-to-image model to visualize an abstract geometric form as a lived space. By undergoing a visual transformation, abstract geometric form becomes architecture. The authors determined the workflow used in the final phase of the design process to convert detailed architectural drawings into renders. As it showed sufficient results, we ought to test whether achieving the same effect using this workflow is possible if the input is not a precise architectural drawing but an image of an abstract geometric form (spatial concept) that does not have architectural characteristics yet.



Figure 3: Example of results made with the workflow to convert detailed sketch into render. The images were created in Dzine software by the authors of this paper.

Input images selection

The first step was to create an abstract geometric form. We chose an image (in further text, “form image”) depicting an abstract geometric form made of orthogonal intertwining white cubes on a light grey background. An image was generated using Midjourney software [10] but could depict a conceptual physical or 3D model. Then, we chose the architecture that we wanted to create. Images of architecture (in further text, “style image”) were found on Pinterest and were deliberately selected to depict certain architectural elements. The first style image shows an object similar to the abstract form. Besides the transparent curtain wall surfaces, subtle greenery, and façade texture, the image looks more or less like an abstract stack of white cubes. The second image has more color and different textures - wooden brisoleils and stone façade, as well as segments of the interior. The third is similar to the second, with slightly different materials and colors. Fourth, a mountain house with a gable roof was intentionally chosen to test whether AI can recognize and implement architectural elements despite the two images’ remarkable differences in form (shape). That question was even more elevated in the subsequent two iterations, where the style images depict surreal architecture with elements of curved geometry. They also vary in perspective and surroundings. Style images gradually digress from the form depicted in the form image, with the first resembling the abstract geometric form and the last significantly different.

Software selection

Previous research has shown that text-to-image software could not reliably represent exact spatial and architectural forms. This leads us to conclude that this process must be carried out in image-to-image-based software with an integrated structure recognition function that differentiates form (structure) from style [7] [12]. We chose Dzine (former Styler.ai) for this research [4]. The image-

Table 1: Results of image-to-image generation

	Result	Style image	Result
Variation 1			
Variation 2			
Variation 3			
Variation 4			
Variation 5			
Variation 6			

to-image model in the Dzine software is based on five input parameters: form image, style image, structure match, style intensity, and text prompt [4]. For this research, all parameters were permanently fixed except the style image. Structure and style match were set to maximum intensity, with no text prompt. The research was conducted in a single iteration for every style image, generating four different result images. The table below shows the best images of the results from every iteration according to the authors.

DISCUSSION

The evaluation of results is conducted through six parameters: materials, lighting, architectural elements (windows, doors, stairs, brisoleil), greenery, surroundings, and overall ambience, and it is shown in the table below. The evaluation was conducted regarding how well the best result of each variation (shown in Table 1) logically and spatially responds to each parameter. Plus means that it responds completely, plus and minus means that it is somewhat logical but not entirely, and the minus means that the result is utterly illogical regarding the parameter. The best score was obtained from Variation 2, where the link between form and style image is quite obvious. The software did not need much imagination to create a white modernistic villa from a stack of white cubes.

Table 2: Evaluation of results

	materials	lightning	windows, doors stairs, brisoleil	geenery	surrounding	overall ambi- ence
Variation 1	+	+	+	-	+ / -	+
Variation 2	+	+	+ / -	+	+	+
Variation 3	+	+ / -	-	-	+ / -	-
Variation 4	+ / -	+	+ / -	+ / -	+	+ / -
Variation 5	+ / -	+	-	/	+	+
Variation 6	+ / -	+	-	+	+	+ / -

An interesting result was created in the last iteration, where the difference between geometries shown in the form and style image was the biggest. Although structure match was set to maximum value, ensuring the consistency of the form, the software produced some new geometry that could serve as an inspiration or change of perspective in further design. In Variation 5, the software recognized windows as architectural elements but did not interpret their shape, only their materialization and lightning. This leads us to conclude that even image-to-image software with integrated structure recognition fails to understand spatial relations completely.



Figure 4: One of the results generated in iteration variation 6.

CONCLUSION

Visualizing an abstract geometric form as architecture using AI software is possible to some extent by using image-to-image-based software with built-in structure recognition. This software recognizes the form shown on the form input image, while the colors and textures are mainly taken from the style input image. Although this shows excellent improvements compared to text-to-image-based software, we cannot say that AI has human-level spatial reasoning skills. This research shows that AI creates the best results where the link between the input and the desired result is obvious. This finding emphasizes that the architect's role as a curator of the process is crucial and that the results largely depend on the carefully chosen input images and parameters. Hence, AI software can be a powerful time-saving tool, but it has a long way to go before replacing the entire creative process. While an architect can see a house, opera, or whole city looking at a crumpled paper, creating this link remains difficult for AI without the exact guidelines.

REFERENCES

- [1] ArchiVinci. Accessed March 28, 2025. <https://www.archivinci.com/>
- [2] Cho, Jaemin; Zala, Abhay; and Bansal, Mohit. 2023. "DALL-Eval: Probing the Reasoning Skills and Social Biases of Text-to-Image Generation Models." In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 3020-3031.
- [3] Conwell, Colin, and Ullman, Tomer D. 2022. "Testing Relational Understanding in Text-Guided Image Generation." In *Computing Research Repository (CoRR)*.
- [4] Dzine. Accessed March 28, 2025. <https://www.dzine.ai>
- [5] Hafri, Alon, and Firestone, Chaz. 2021. "The perception of relations." In *Trends in Cognitive Sciences*, 25(6), 475-492.
- [6] Horvath, Anca-Simona, and Pouliou, Panagiota. 2024. "AI for conceptual architecture: Reflections on designing with text-to-text, text-to-image, and image-to-image generators." In *Frontiers of Architectural Research*, 13, 593-612.
- [7] Karadag Ilker. 2023. "Transforming Sketches into Realistic Images: Leveraging Machine Learning and Image Processing for Enhanced Architectural Visualization." In *Sakarya University Journal of Science*, 27(6), 1209-1216.
- [8] Lee Tony; Yasunaga, Michihiro, Meng, Chenlin; Mai, Yifan; Park, Joon Sung; Gupta, Agrim; Zhang, Yunzhi; Narayanan, Deepak; Teufel, Hannah Benita; Bellagente, Marco; Kang, Minguk; Park, Taesung; Leskovec, Jure; Zhu, Jun-Yan; Fei-Fei, Li; Wu, Jiajun; Ermon, Stefano; and Liang, Percy. 2023. "Holistic Evaluation of Text-to-Image Models." In arXiv.org.
- [9] Marcus, Gary; Davis, Ernest; and Aaronson, Scott.. 2022. "A very preliminary analysis of DALL-E 2." In arXiv.org.
- [10] Midjourney. Accessed March 28, 2025. <https://www.midjourney.com/home>
- [11] Oppenlaender, Jonas. 2022. "The Creativity of Text-based Generative Art." In arXiv.org.
- [12] PromeAI. Accessed March 28, 2025. <https://www.promeai.pro>
- [13] Rezwana, Jeba, and Maher, Mary Lou. 2022. "Designing Creative AI Partners with COFI: A Framework for Modeling Interaction in Human-AI Co-Creative Systems." In arXiv.org.
- [14] Stigsen, Mathias Bank; Moisi, Alexandra; Rasoulzadeh, Shervin; Schinegger, Kristiona; and Routzinger Stefan. 2023. "AI Diffusion as Design Vocabulary - Investigating the use of AI image generation in early architectural design and education." In eCAADe conference.
- [15] Sukkar, Ahmad W.; Fareed, Mohamed W.; Yahia, Moohammed Wasim; Abdalla, Salem Buhashima; Ibrahim, Iman; and Senjab, Khaldoun Abdul Karim. 2024. "Analytical Evaluation of Midjourney Architectural Virtual Lab: Defining Major Current Limits in AI-Generated Representations of Islamic Architectural Heritage." In *Buildings*, 14, 786-811.