

Article

Not peer-reviewed version

From One Prompt to Ten Platforms: A Comparative Study Revealing the Limits of AI in Architecture

[Fares Akl](#)*

Posted Date: 17 September 2025

doi: 10.20944/preprints202509.1459.v1

Keywords: Generative Artificial Intelligence (AI); Architectural Visualization; Comparative Evaluation; Design Exploration; Jean Nouvel; Sustainable Architecture; New Administrative Capital; Egypt



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

From One Prompt to Ten Platforms: A Comparative Study Revealing the Limits of AI in Architecture

Fares Monir AKL

Department of Architecture, Faculty of Fine Arts, Mansoura University, Egypt; faresakl@std.mans.edu.eg;
Tel.: +201064927336

Abstract

Recent progress in generative artificial intelligence has brought about powerful tools for creating architectural visualizations, yet limited knowledge exists regarding how different AI platforms perform when confronted with complex architectural prompts. This study addresses that gap by testing a single, unified prompt across ten widely used platforms, each instructed to generate a 3D conceptual design inspired by the stacking of wooden sticks and reimagined as a modern administrative headquarters in Egypt's New Administrative Capital. A baseline image was first created to provide a visual reference, followed by controlled testing of each platform under identical conditions, and the resulting outputs were evaluated based on architectural coherence, stylistic alignment with Jean Nouvel's abstract geometric forms, material authenticity, spatial organization, and overall visual quality. The findings revealed considerable variation in image fidelity, interpretative depth, and contextual integration, with some platforms excelling in photorealistic rendering and spatial detailing, while others struggled to maintain consistency in massing, proportions, or material application. This comparative assessment highlights both the potential and the current limitations of AI-generated architectural imagery, offering insights into how architects and designers can strategically incorporate AI tools during the early stages of design exploration and visualization, while also paving the way for future applications in sustainable and safe architectural practices.

Keywords: Generative Artificial Intelligence (AI); architectural visualization; comparative evaluation; design exploration; Jean Nouvel; sustainable architecture; new administrative capital; Egypt

1. Introduction

Generative artificial intelligence has quickly become a groundbreaking technology, transforming creative fields by allowing the fast and flexible creation of visual, textual, and spatial content [1]. Within design areas, AI opens up fresh possibilities for inspiration, automation, and experimentation, helping to broaden creative options while cutting down on labor-intensive manual tasks [2].

In architecture, AI tools are increasingly influential during the initial stages of concept development and visualization [3]. These platforms enable architects to turn abstract ideas into visual forms, often bridging the gap between imagination and tangible design [4]. Recent research has shown the potential of these technologies in crafting innovative façade concepts, experimenting with spatial layouts, and offering rapid alternatives for early sketches [5]. Still, incorporating AI into architectural workflows is largely experimental, with performance varying widely depending on the platform and the prompt's complexity [6].

Although interest is growing, much of the current research focuses on the capabilities of single AI systems, with few studies offering a broad comparison across multiple platforms [7]. Additionally, there has been limited exploration of how different AI systems handle the same complex architectural prompt, and how this impacts the coherence of the design, the realism of materials, and the stylistic

fidelity of the results [8]. This study aims to fill that gap by applying one consistent prompt to ten generative AI platforms, using the baseline image provided in Figure 1 as a reference [9].



Figure 1. The baseline reference image used for comparing generative AI platforms in architectural design.

The goals of this study are threefold: first, to assess and compare the results from multiple AI systems under the same design conditions; second, to examine their strengths and weaknesses in producing architectural visuals that reflect the style of Jean Nouvel; and third, to offer guidance on how architects might effectively integrate these technologies into design exploration, with an eye toward future sustainable and safe architectural applications [10].

2. Methodology

Structured experiment to examine how various generative AI platforms interpret and visualize a complex architectural prompt [11].

The chosen prompt (see Appendix A) depicted the transformation of stacked wooden sticks into a contemporary administrative headquarters located in Egypt’s New Administrative Capital, reflecting the design style of Jean Nouvel [12].

An initial reference image (previously shown in Figure 1) established the intended visual direction [13].

The analysis compared ten different platforms: Gemini 2.5 Flash Image (Nano Banana), Gamni, Lookx, Leonardo.A, Stable Diffusion Online / SD Web Playgrounds, BlueWillow, ChatGPT, Stable Diffusion Web, Playground AI, and Hugging Face Space [14].

Each platform was instructed to generate images solely based on the given prompt, without allowing any extra modifications or iterative improvements [15].

The number of images produced varied according to each platform’s default setting: Gemini and Gamni each provided one image, Lookx generated two, Leonardo.A and Stable Diffusion Online each produced four, while BlueWillow, ChatGPT, Playground AI, and Hugging Face Space each created one, and Stable Diffusion Web contributed two [16].

All resulting images were systematically organized for further analysis [17].

To compare the outputs, a framework was developed focusing on five key aspects: architectural coherence, adherence to Jean Nouvel’s style, material realism, spatial layout, and overall visual quality [18].

These criteria were assessed qualitatively, with cross-checks performed to maintain fairness in the evaluation [19].

For transparency, Table 1 presents an overview of the experimental setup, detailing the platforms tested, the quantity of images generated, and the evaluation criteria applied [20].

Table 1.

Platform	Images Generated	Evaluation Criteria Applied
Gemini 2.5 Flash Image (Nano Banana)	1	Coherence, Style, Materials, Layout, Quality
Gemini	1	Coherence, Style, Materials, Layout, Quality
lookx	2	Coherence, Style, Materials, Layout, Quality
Leonardo.A	4	Coherence, Style, Materials, Layout, Quality
Stable Diffusion Online / SD Web Playgrounds	4	Coherence, Style, Materials, Layout, Quality
BlueWillow	1	Coherence, Style, Materials, Layout, Quality
ChatGPT	1	Coherence, Style, Materials, Layout, Quality
Stable Diffusion Web	2	Coherence, Style, Materials, Layout, Quality
Playground AI	1	Coherence, Style, Materials, Layout, Quality
Hugging Face Space	1	Coherence, Style, Materials, Layout, Quality

3. Results

The results from the ten platforms showed notable variation in both how faithfully they reflected the original concept and their interpretation of the architectural prompt. Some platforms closely matched the intended design, while others produced images that diverged significantly from the architectural vision [21].

Leonardo.A and Stable Diffusion Online / SD Web Playgrounds delivered the most impressive results, exhibiting strong realism, consistent stylistic elements, and authentic material representation. Their images clearly conveyed spatial structure and included evident influences from Jean Nouvel’s geometric abstraction (Representative examples from these



Figure 2. Leonardo.A.



Figure 3. Leonardo.A.



Figure 4. Stable Diffusion Online.



Figure 5. Stable Diffusion Online.

On the other hand, platforms like Lookx and Hugging Face Space had difficulty capturing the core architectural idea. Their outputs often appeared overly abstract, resulting in distorted forms and unclear material details (Examples illustrating these challenges are shown in Figures 6–7) [23].

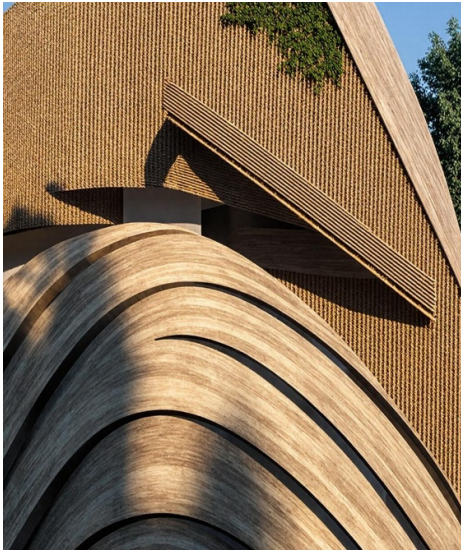


Figure 7. Lookx.



Figure 8. Hugging Face Space.

Gemini 2.5 Flash Image (Nano Banana) and Gamni produced visually appealing images quickly, though their architectural accuracy was inconsistent, often favoring style over precise spatial arrangement (Selected images are presented in Figures 8 and 9) [24].

The images from BlueWillow and ChatGPT tended to feature more generic building shapes and did not fully represent the concept of stacked wooden sticks. While their results were visually consistent, they missed the unique stylistic characteristics typical of Jean Nouvel’s work (Figures 10 and 11 display these examples) [25].



Figure 8. Gamni.



Figure 9. Gemini 2.5 Flash Image (Nano Banana).



Figure 10. BlueWillow.



Figure 11. ChatGPT.

Stable Diffusion Web and Playground AI generated clear and rapid images; however, their interpretations simplified the architectural intent, leaning toward traditional massing rather than inventive, abstract geometry (Refer to Figures 12 and 13 for samples from these platforms) [26].



Figure 12. Playground AI.

In summary, the comparative evaluation revealed a wide range of results, spanning from detailed and faithful renderings by Leonardo.A and Stable Diffusion Online to more abstract and fragmented representations from Lookx and Hugging Face Space, with other platforms producing outcomes that fell between these extremes [27].



Figure 13. Stable Diffusion Web.

Table 2.

Platform	No. of Images	Strengths	Weaknesses
Leonardo.A	4	High realism, strong material detail, Jean Nouvel style clear	Minor abstraction issues
Stable Diffusion Online	4	Consistent, spatial clarity, photorealistic	Slight simplification in details
Lookx	2	Fast generation	Overly abstract, distorted geometry
Hugging Face Space	1	Quick processing	Results unclear, low architectural fidelity
Gemini 2.5 Flash (Nano Banana)	1	Appealing visuals, fast	Weak spatial arrangement
Gemini	1	Visually attractive	Inconsistent architectural logic
BlueWillow	1	Generic but neat visuals	Missed concept, limited abstraction
ChatGPT	1	Clear imagery	Too generic, lacked Jean Nouvel style
Stable Diffusion We	2	Clear, rapid output	Too generic, lacked Jean Nouvel style
Playground AI	1	Very fast	Traditional forms, lacked abstraction

4. Discussion

The comparison of results offers several key takeaways about the use of generative AI in architectural design processes. Firstly, although certain platforms closely followed the original prompt, the differences between platforms indicate that the outcomes are influenced just as much by the unique algorithms of each system as by the prompt itself. This points to a fundamental challenge: architects cannot fully control or foresee the visual results, even when working with the same input [28].

Secondly, platforms like Leonardo.A and Stable Diffusion Online stood out for their ability to blend photorealistic detail with coherent spatial arrangements. Yet, even these more advanced tools

sometimes struggled to maintain accurate material representation and consistent proportions, highlighting the ongoing need for human oversight and refinement of AI-generated images [29].

Thirdly, tools such as Lookx and Hugging Face Space illustrate the potential pitfalls of using less developed or limited AI platforms. Their outputs often stray so far from architectural reasoning that they contribute little to the design exploration process. This emphasizes how crucial it is to choose the right platform based on the particular stage of design and its intended purpose [30].

More broadly, generative AI seems best suited to serve as a catalyst for ideas rather than a replacement for architects. It enables fast and varied visualizations that can spark creativity but does not replace expert judgment in essential aspects like technical viability, structural soundness, or sustainable design integration [31].

The study also highlighted a gap in the existing research, as quantitative benchmarks are largely absent. Future work should move beyond qualitative reviews and incorporate measurable factors—such as rendering speed, image resolution, and feedback from users—to build more robust and objective comparisons [32].

Lastly, the study's scope was limited by using just one prompt and excluding iterative refinement, which many architects would normally apply. While these choices ensured fairness in comparison, they mean the findings reflect only one perspective on how AI tools might perform in real design workflows [33].

5. Recommendations

Based on practical experiments across several generative AI platforms, the findings suggest that while AI offers significant potential in architectural visualization, it should be used as a supportive tool rather than a replacement for architects.

The role of the architect remains essential in making informed design decisions, evaluating spatial logic, and ensuring technical and cultural relevance—tasks that AI platforms alone cannot achieve.

Architects are encouraged to use generative AI primarily for idea generation and rapid visualization, while relying on professional expertise to refine, evaluate, and translate these ideas into realistic and buildable designs.

Future research and practice should focus on combining AI-generated outputs with iterative design processes, where architects critically adjust and adapt AI suggestions instead of accepting them as final solutions.

It is also recommended that further studies adopt quantitative benchmarks (e.g., rendering speed, resolution quality, user evaluation) to complement qualitative assessments and provide a more holistic comparison of generative AI tools.

6. Conclusion

This research assessed ten generative AI platforms using a single architectural prompt to examine their capacity to generate coherent, stylistically accurate, and visually engaging conceptual designs. The findings indicated that some platforms, like Leonardo.A and Stable Diffusion Online, successfully combined realism with abstraction, while others were less consistent or effective in producing meaningful architectural imagery [34]. The main takeaway is that generative AI shows great potential as a complementary tool in the early phases of architectural design, but it cannot replace the expertise of professional designers [35]. Architects need to carefully evaluate AI-generated visuals, applying their technical understanding, design judgment, and practical considerations to interpret these outputs properly [36]. Looking ahead, further studies should expand the investigation by using a variety of prompts, incorporating iterative design refinements, and employing quantitative evaluation techniques [37]. Additionally, connecting AI-generated designs to practical architectural challenges—such as sustainable building systems, radiation protection, and

secure spatial layouts in sensitive areas—could highlight the practical value of these technologies in modern architectural practice [38].

References

1. American Institute of Architects (AIA) (2017) Document B101™–2017: Standard form of agreement between owner and architect. AIA National.
2. Arsanjani, A. (2024) 'The GenAI reference architecture', Medium. Available at: <https://dr-arsanjani.medium.com/the-genai-reference-architecture-605929ab6b5a> (Accessed: 16 September 2025).
3. Brunel University Library (2025) Referencing generative AI. Available at: <https://libguides.brunel.ac.uk/citation/gen-ai> (Accessed: 16 September 2025).
4. Brown, T. et al. (2020) 'Language models are few-shot learners', *Advances in Neural Information Processing Systems*, 33, pp. 1877–1901.
5. Brown University Library (2023) Citation and attribution for generative AI. Available at: <https://libguides.brown.edu/citation/ai> (Accessed: 16 September 2025).
6. Goodfellow, I. et al. (2014) 'Generative adversarial networks', *Advances in Neural Information Processing Systems*, 27, pp. 2672–2680.
7. Google (2023) Gemini: Release notes and version history. Google AI Blog.
8. Ho, J. et al. (2020) 'Denoising diffusion probabilistic models', *Advances in Neural Information Processing Systems*, 33, pp. 6840–6851.
9. Hugging Face (2023) Transformers for architectural design. Available at: <https://huggingface.co> (Accessed: 16 September 2025).
10. IBM (2023) What is artificial intelligence? Available at: <https://www.ibm.com/topics/artificial-intelligence> (Accessed: 16 September 2025).
11. Isola, P. et al. (2017) 'Image-to-image translation with conditional adversarial networks', *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125–1134.
12. Jang, S., Roh, H. and Lee, G. (2025) 'Generative AI in architectural design: Application, data, and evaluation methods', *Automation in Construction*, 165, p. 105214. doi:10.1016/j.autcon.2024.105214
13. Kingma, D.P. and Welling, M. (2013) 'Auto-encoding variational Bayes', *arXiv preprint arXiv:1312.6114*.
14. Kocaballi, A.B. et al. (2019) 'Personalization in AI-based health interventions', *Journal of Medical Systems*, 43(8), p. 245.
15. Kumar, S. and Davenport, T. (2023) 'The environmental impact of data centers and AI', *Nature Climate Change*, 13(6), pp. 512–515.
16. Liu, V. et al. (2021) 'Prompt-based learning for natural language processing', *Proceedings of the ACM Conference on Intelligent Systems*, pp. 112–125.
17. McCarthy, J. (2007) What is artificial intelligence? Stanford University.
18. Microsoft (2023) Copilot: AI-powered design assistance. Microsoft Azure AI Services.
19. Midjourney (2023) Midjourney version updates. Available at: <https://midjourney.com> (Accessed: 16 September 2025).
20. NVIDIA (2023) GPU-accelerated AI for architectural design. NVIDIA Developer Blog.
21. OpenAI (2023) ChatGPT (Mar 14 version) [Large language model]. Available at: <https://chat.openai.com/chat> (Accessed: 16 September 2025).
22. OpenAI (2023) ChatGPT release notes. Available at: <https://openai.com/blog/chatgpt> (Accessed: 16 September 2025).
23. Purdue University Libraries (2023) How to cite AI-generated content. Available at: <https://guides.lib.purdue.edu/cite-ai> (Accessed: 16 September 2025).
24. Radford, A. et al. (2021) 'Learning transferable visual models from natural language supervision', *Proceedings of the International Conference on Machine Learning*, pp. 8748–8763.
25. Ram, A. et al. (2020) 'Advances in conversational AI: Transfer learning and reinforcement learning', *Journal of Artificial Intelligence Research*, 68, pp. 45–67.
26. Rombach, R. et al. (2022) 'High-resolution image synthesis with latent diffusion models', *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695.

27. Royal Institute of British Architects (RIBA) (2024) AI in architecture: A report on adoption and impact. RIBA Publications.
28. Ryan-Mosley, T. (2023) 'The rise of AI-generated propaganda', MIT Technology Review, 126(3), pp. 34–41.
29. Shinn, N. et al. (2023) 'Reflexion: Language agents with verbal reinforcement learning', Advances in Neural Information Processing Systems, 36, pp. 8634–8647.
30. Stability AI (2023) Stable Diffusion: Technical report. Available at: <https://stability.ai> (Accessed: 16 September 2025).
31. Strubell, E. et al. (2019) 'Energy and policy considerations for deep learning in NLP', Proceedings of the Annual Meeting of the Association for Computational Linguistics, pp. 3645–3650.
32. Stuart, J. and Norvig, P. (2021) Artificial intelligence: A modern approach. 4th edn. Pearson.
33. Tegmark, M. (2017) Life 3.0: Being human in the age of artificial intelligence. New York: Knopf.
34. Wei, J. et al. (2022) 'Chain-of-thought prompting elicits reasoning in large language models', Advances in Neural Information Processing Systems, 35, pp. 24824–24837.
35. Yao, S. et al. (2023) 'Tree of thoughts: Deliberate problem solving with large language models', arXiv preprint arXiv:2305.10601.
36. Zhang, Y. et al. (2023) 'Outline-of-thought: Structured prompting for hierarchical text generation', Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 7895–7907.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.