Spot the Bots: Analyzing Text-to-Image Outputs for the Field of Landscape Architecture

Emily Schlickman[1], Xinyi Li[2], Danxiang Wang[2]

[1]University of California, Davis, California/USA · eschlickman@ucdavis.edu
[2]Harvard University, Massachusetts/USA

Abstract: Text-to-image generation is a design tool that employs artificial intelligence to create imagery from natural language prompting. While the use of text-to-image platforms is gaining momentum in design fields and has the potential to disrupt non-generative design processes, there is limited analysis of the technology in relation to landscape architecture and allied fields. Thus, this paper explores how text-to-image platforms can be used in the design process, what its current limitations are, how its outputs compare to those produced without artificial intelligence, and if people can identify machine-generated outputs. To do this, the research team tested three text-to-image platforms, created a series of renderings from the three platforms using the same prompt, and found similar images generated without artificial intelligence. The team then set up an exercise to see if people could correctly identify which images had been machine-generated to better understand the current capabilities of the programs. In analyzing the results from the exercise, the team found that participants had a difficult time distinguishing between machine-generated images and human-generated images as only 43% of the guesses were correct. The team also found that while text-to-image tools have the greatest potential when used on the front-end of a project when quick ideation and iteration is key, the tools currently face a number of limitations including reductionism and inappropriate detail capture.

Keywords: Artificial intelligence, text-to-image generation, generative design, landscape architecture

1 Introduction

Generative design, which employs artificial intelligence algorithms to develop design alternatives from guided inputs, is increasingly being employed within landscape architecture and its allied fields of architecture, urban design, and urban planning to tackle increasingly complex design challenges. While originally used to develop new patterns and forms for architectural elements like facades (BUHAMDAN et al. 2020), generative design is now being used for a number of applications including floor plan generation, building performance optimization and construction facilitation (YILDIRIM 2022, OZEROL & SELCUK 2023).

One subset of generative design currently rising in popularity is text-to-image generation which involves natural language prompting to create novel imagery. While text-to-image platforms have the potential to disrupt non-generative design processes through enhanced visualization capabilities, concept development, prototyping, collaboration, and accessibility opportunities, there is limited analysis of the technology as it relates to landscape architecture.

This paper unpacks a recent research project entitled "Spot the Bots" to provide an analysis of current text-to-image technology for the field of landscape architecture. Four primary research questions drove the project: 1 – How might text-to-image tools be employed for the field of landscape architecture? 2 – What are the current limitations to widescale adoption and use? 3 – How do machine-generated images compare to human-generated images? And 4 – Can people distinguish between the two sets of images?

Journal of Digital Landscape Architecture, 9-2024, pp. 764-771. © Wichmann Verlag, VDE VERLAG GMBH · Berlin · Offenbach. ISBN 978-3-87907-752-6, ISSN 2367-4253, e-ISSN 2511-624X, doi:10.14627/537752071.

2 Methodology

To answer these questions, our research team tested three text-to-image platforms– DALL-E 2, Midjourney, and Stable Diffusion.[1] The team organically explored each platform to better understand user interfaces and capabilities before creating a series of images using the same, sequentially-constructed prompt. This prompt was developed based on basic prompt engineering protocols using descriptors commonly used to describe landscape renderings: "From a bird's eye perspective, the scenic view of a futuristic modern city, with utopian eco-friendly landscape design of wetlands, with water and a bridge, many pocket wetlands in the middle, near the Sacramento River, sunset light, hyperdetailed, ultra-realistic".[2]

After generating imagery from the three text-to-image platforms, the research team found human-generated imagery created by design firms. The goal was to find a set of machine-generated and human-generated images that generally reflected the prompt above.

The research team then developed the "Spot the Bots" exercise by curating 30 images – 18 human-generated images and 12 machine-generated images using one of the three platforms above. Once selected, all 30 images were formatted and printed in the same way (8x8" squares) and pinned up on a cork board in the lobby of an academic building that houses a landscape architecture and environmental design program. The machine-generated images were intentionally mixed up with human-generative images into two areas – on the left, nine images were machine-generated and on the right, three were machine-generated. The directions, posted in between the two areas, asked people to grab a few stickers from a provided roll and to put the stickers under the images they believed to have been generated by machines (Fig. 1). The research team then emailed students, staff members, and faculty members in the landscape architecture and environmental design program to encourage participation.

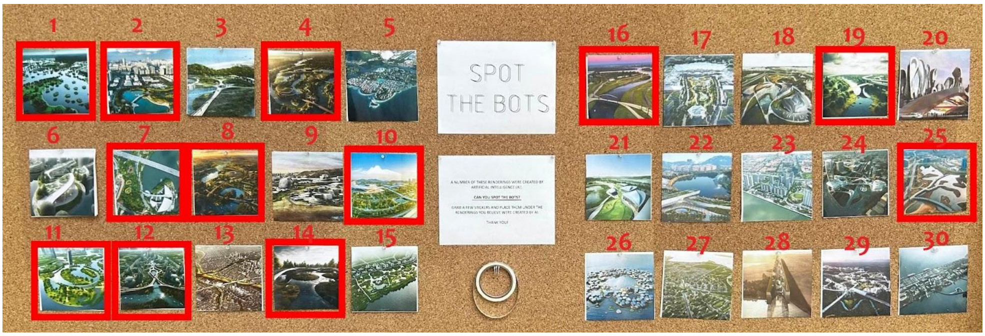

Fig. 1: The array of images used for the "Spot the Bots" exercise, with AI-generated images outlined in red (the red notations were not part of the visual display but added afterwards for explanatory purposes).

[1] Please note that we conducted this study in the spring of 2023, so we were limited to the tools and versions that were available then. While there are many other platforms including DALLE-3, Stable Diffusion pretrained, Stable Diffusion pretrained XL (SDXL V1.0), testing these was outside the scope of this project.

[2] It should also be noted that while some of the platforms had advanced features (like inpainting and outpainting) at the time of testing, none were used in order to do a direct comparison between the three programs.

3 Findings

Over the course of one week during the spring quarter of 2023, 361 votes were made as to which images had been produced by text-to-image platforms, and of those 361 votes, 155 (or 43%) were correct (Fig. 2). Of the ten images with the most dots (most votes for being machine-generated), four were produced by text-to-image platforms and of the ten images with the least dots (least votes for being machine-generated), two were produced by text-to-image platforms (Fig. 3).

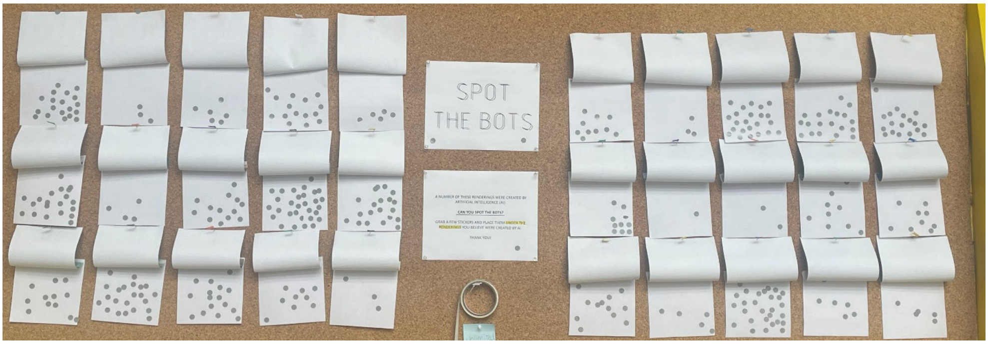

Fig. 2: Raw results from the "Spot the Bots" exercise, with dots representing votes for images thought to be machine-generated

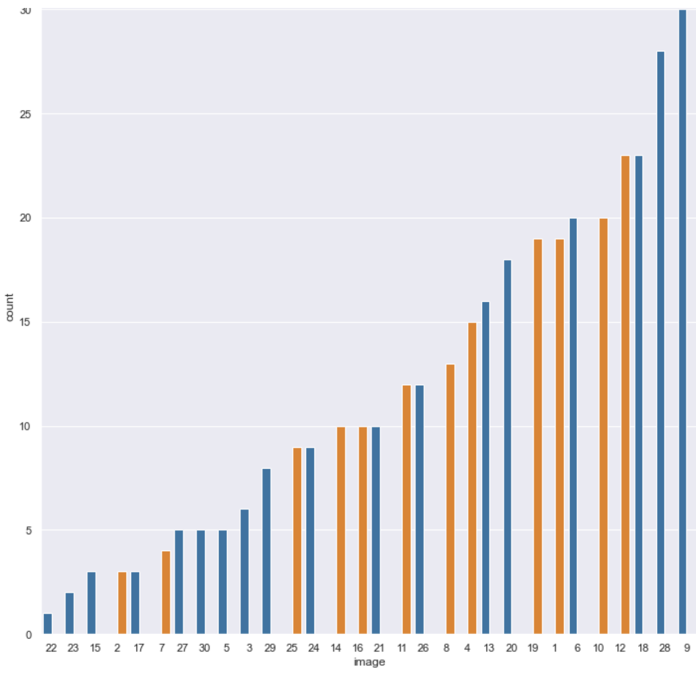

Fig. 3: Data showing number of votes (for being machine-generated) per image. Orange lines indicate machine-generated imagery.

In analyzing the results from the "Spot the Bots" exercise, the research team realized that participants had a difficult time distinguishing between the machine-generated imagery and human-generated imagery, as less than half of the guesses for "Spot the Bots" were correct. This is important because it reveals that programs like DALL E2, Midjourney, and Stable Diffusion are already capable of creating realistic architectural renderings and, thus, have the potential to replace or augment non-AI work that designers are already doing. Unfortunately, though, since the research team did not have the opportunity to interview participants to better understand their voting logic, they could only speculate about the reasons for this. One speculation is that participants might have assumed that visually impressive images with intricately woven roads and bridges and uniquely shaped buildings (Figure 4) were machine-generated due to the complexity of the forms. If so, this might mean that participants tended to overestimate the capacity of the text-to-image programs. Another speculation is that participants might have assumed that images showing in-progress (or under construction) design work were human-generated. Two examples of this are shown in Figure 5 – both, of which, were machine-generated and received relatively few votes. One final speculation relates to atmospheric qualities of the imagery – perhaps participants were looking for particularly advanced lighting, shadow, and cloud effects (Fig. 6) when searching for machine-generated images.

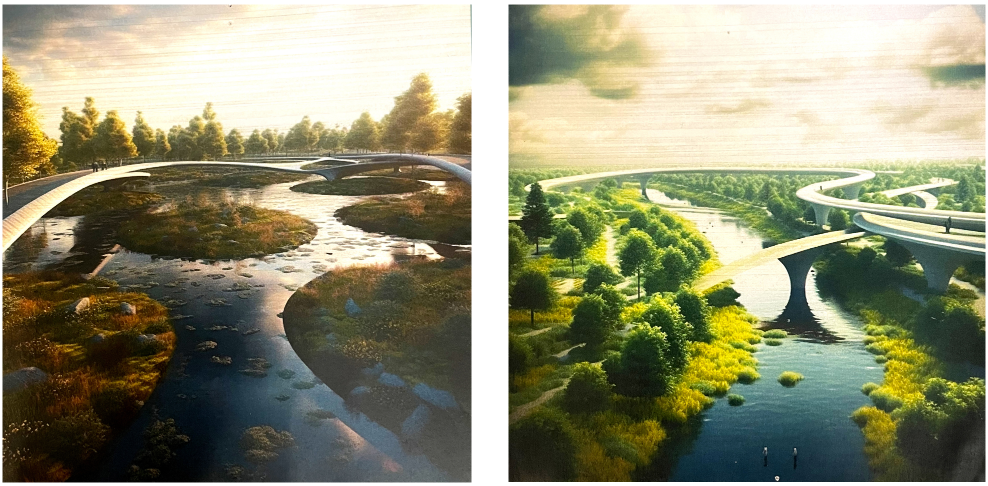

Fig. 4: On the left is image #19 and on the right is image #14. Both are machine-generated and feature intricately woven roads and bridges.

Fig. 5: On the left is image #25 and on the right is image #2. Both are machine-generated and appear to be showing sites under construction.

Fig. 6: On the left is image #4 and on the right is image 82. Both are machine-generated and feature advanced lighting, shadow and cloud effects.

The research team also determined that text-to-image tools could provide the greatest value when employed on the "fuzzy front end" (KOEN et al. 2001) of a design project when quick ideation is necessary. In under one minute, all three platforms provided a range of visual outputs, which the team thought could be built upon and modified in later design stages (PANICKER 2022). This act of co-creation to expand early design ideas has been likened to hive intelligence (CANTRELL & ZHANG 2021).

That being said, the team also unpacked a range of limitations with the platforms that might discourage widescale adoption and use in the field of landscape architecture. One limitation is that the outputs from these platforms are only as good as the inputs (prompts) and there is little guidance available for how to effectively construct a prompt (OPPENLAENDER 2022). These platforms also struggle with complex design problems and have been labelled as being overly reductionist – focusing more on what projects could look like instead of how they could perform (BOLOJAN et al. 2022). Additionally, it is often difficult for the platforms to create realistic contexts for imagery as they do not recognize geographic locations (SENEVIRATNE et al. 2022) and for platforms to create practical design solutions. Appropriate detail capture is also a challenge as many outputs have random blurred areas or incorrect material overlays. The research team also found themselves aimlessly iterating and relying too heavily on the tools for ideation and project development (PAANENEN et al. 2023). Lastly, the team found the generation/training process to be opaque and could lead to biased outputs with harmful and incorrect content as well as outputs with potential copyright issues (OPPENLAENDER 2022).

Additionally, the team analyzed the limitations of their own study to reveal shortcomings and new possible directions for future research. To begin, the research team likely had some bias when selecting what imagery to use for the "Spot the Bots" exercise – both in selecting the machine-generated imagery and the human-generated imagery. If replicated in the future, it would be helpful to embed some sort of randomization into this selection process. Furthermore, there might have been some participant bias during the voting process as some people might have been deterred or encouraged by the number of stickers they viewed when lifting up different images on the board. Perhaps, in the future, previous votes could be obscured so as to not sway future participants. Participants might have also recognized some of the human-generated imagery depending on their experience in the field and their knowledge of design firms. Also, since "Spot the Bots" was set up in a public building entry, it was difficult to get a sense for who voted on the images and how many stickers each person used; thus, monitoring or recording participation and limiting the number of votes per person could be future considerations to ensure that the sample is representative of the larger landscape architecture community. Additionally, the exercise was only up for one week at the end of the quarter – right before summer vacation – perhaps if it were launched in the middle of the academic year and if it were up for longer, more people would have participated, generating a wider pool of responses. Lastly, while the research team has speculated about why participants had a difficult time distinguishing between machine-generated imagery and human-generated imagery, it could be helpful to interview or survey participants to better understand their selection process and their perspectives on the efficacy of certain images in selling design concepts, demonstrating impact, etc. Alternatively, the research team could put together a follow-up "Spot the Bots" exercise and intentionally select machine-generated images they believe to be less detectable based on findings from this study.

4 Conclusion and Outlook

The "Spot the Bots" exercise outlined in this paper revealed that relatively nascent generative design tools like DALL E2, Midjourney, and Stable Diffusion are already succeeding at creating realistic architectural renderings – so much so that they are even capable of tricking designers with discerning eyes. And all of these tools are changing and improving so quickly given the rapid rate of development with AI technology.

So, what does this mean for the field of landscape architecture? How might AI tools change the design process? Is it possible that these tools are an existential threat to our discipline?

Some believe this could be the case. In fact, in March 2023, Goldman Sachs released a report stating that generative artificial intelligence could affect 300 million full-time jobs, with 37% of these being in the fields of architecture and engineering (GOLDMAN SACHS 2023). And while this may, indeed, become a reality, our research team came out of this project with a more nuanced outlook about the current role of AI in design. Given our analysis, we do not believe that current text-to-image platforms can replace the detailed, specialized, and functional work that designers do on a daily basis. Rather, we believe that these tools could actually help designers work more efficiently and effectively, and could shift our role from being sole authors to co-authors in the production of more hybridized design work. This is echoed in the words of David Holz, founder of Midjourney: “The goal is to make humans more imaginative, not make imaginative machines, which I think is an important distinction” (SALKOWITZ 2022).

As we embark into a world with increasingly challenging design problems, generative artificial intelligence tools have the potential to disrupt, restructure, and reshape creative disciplines. Thus, it is critical for landscape architects and designers to test them, critically evaluate their use in the field, and reflect upon their ever-changing role in design.

References

BOLOJAN, D, VERMISSO, E. & YOUSIF, S. (2022), Is Language all we need? A query into architectural semantics using a multimodal generative workflow. Proceedings of the 27th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), 1, 353-362.

BUHAMDAN, S., ALWISY, A. & BOUFERGUENE, A. (2020), Generative Systems in the architecture, engineering, and construction industry: A systematic review and analysis. International Journal of Architectural Computing, 19 (3), 226-249.

CANTRELL, B. & ZHANG, Z. (2021), Artificial Intelligence and machine learning in landscape architecture. In: The Routledge Companion to Artificial Intelligence in Architeture, 232-247.

GOLDMAN SACHS. (2023). Global Economics Analyst: The Potentially Large Effects of Artificial Intelligence on Economic Growth. https://www.gspublishing.com/content/research/en/reports/2023/03/27/d64e052b-0f6e-45d7-967b-d7be35fabd16.html (20.09.2023).

KOEN, P., AJAMIAN, G., BURKART, R., CLAMEN A., DAVIDSON, J., D’AMORE, R., ELKINS, C. et al. (2001), Providing clarity and a common language to the “fuzzy front end”. Research-Technology Management, 44 (2), 46-55.

OPPENLAENDER, J. (2022), A Taxonomy of Prompt Modifiers for Text-To-Image Generation. ARXIV, https://doi.org/10.48550/arXiv.2204.13988.

OPPENLAENDER, J. (2022), The Creativity of Text-to-Image Generation. Academic Mindtrek, 2022, 192-202.

OZEROL, G. & SELCUK, S. (2023), Machine learning in the discipline of architecture: a review on the research trends between 2014 and 2020. International Journal of Architectural Computing, 21 (1), 23-41.

PAANENEN, V., OPPENLAENDER, J. & VISURI, A. (2023), Using text-to-image generation for architectural design ideation. ARXIV, https://doi.org/10.48550/arXiv.2304.10182.

PANICKER, S. (2022), AI-Inflected Art/Architecture: Who (or rather, what) is the artist/architect? BLUEPRINT SEPTEMBER 2022, 3 (2), 15-36.

SALKOWITZ, R. (2022), Midjourney Founder David Holz On The Impact Of AI On Art, Imagination And The Creative Economy. Forbes. https://www.forbes.com/sites/robsalkowitz/2022/09/16/midjourney-founder-david-holz-on-the-impact-of-ai-on-art-imagination-and-the-creative-economy (20.09.2023).

SENEVIRATNE, S., SENANAYAKE, D., RASNAYAKA, S., VIDANAARACHCHI, R. & THOMPSON, J. (2022), DALLE-URBAN: Capturing the urban design expertise of large text to image transformers. ARXIV, https://doi.org/10.48550/arXiv.2208.04139.

WIJNANDS, J., NICE, K., THOMPSON, J., ZHAO, H. & STEVENSON, M. (2019), Streetscape augmentation using generative adversarial networks: Insights related to health and well-being. Sustainable Cities and Society, 49. https://doi.org/10.1016/j.scs.2019.101602.

YILDIRIM, E. (2022), Text-to-Image Generation A.I. in Architecture. In: Art and Architecture: Theory, Practice and Experience, 97-120.