

Testing the Capability of AI Art Tools for Urban Design

Connor Phillips , Junfeng Jiao , and Emmalee Clubb , University of Texas at Austin, Austin, TX, 78712, USA

This study aimed to evaluate the performance of three artificial intelligence (AI) image synthesis models, Dall-E 2, Stable Diffusion, and Midjourney, in generating urban design imagery based on scene descriptions. A total of 240 images were generated and evaluated by two independent professional evaluators using an adapted sensibleness and specificity average metric. The results showed significant differences between the three AI models, as well as differing scores across urban scenes, suggesting that some projects and design elements may be more challenging for AI art generators to represent visually. Analysis of individual design elements showed high accuracy in common features like skyscrapers and lawns, but less frequency in depicting unique elements such as sculptures and transit stops. AI-generated urban designs have potential applications in the early stages of exploration when rapid ideation and visual brainstorming are key. Future research could broaden the style range and include more diverse evaluative metrics. The study aims to guide the development of AI models for more nuanced and inclusive urban design applications, enhancing tools for architects and urban planners.

Urbanization, a defining phenomenon of the modern era, has given rise to a global trend of rapidly expanding cities, as populations increasingly migrate from rural to urban areas in search of improved economic opportunities and enhanced quality of life. This influx of people into urban environments has strained existing infrastructure, leading to various challenges such as overcrowding, inadequate housing, and limited access to essential amenities. Consequently, the need to redesign and repurpose forgotten or underutilized spaces within these urban landscapes has become paramount to addressing these challenges and ensuring sustainable development. In the face of these urbanization-induced challenges, neighborhood redesign initiatives have traditionally been driven by developers or city governments who recognize the potential for transforming these underutilized spaces into vibrant, functional, and aesthetically pleasing environments. Oftentimes, developers work in collaboration with a

diverse team of professionals, including urban designers, landscape architects, and contractors, to shape these spaces.

Urban designers and landscape architects play a critical role in designing built environments and creating a comprehensive vision for the neighborhood by considering numerous elements, such as land use, transportation, public space, and housing, into a cohesive and functional whole. Furthermore, oftentimes professionals are tasked with creating aesthetically pleasing and environmentally responsible outdoor spaces that promote human well-being, enhance ecological function, and foster a sense of community. Artificial intelligence (AI) image synthesis models are cutting-edge tools that leverage the power of machine learning to generate creative and visually compelling artwork. Image diffusion models, as opposed to generative adversarial networks (GANs), function by gradually transforming a random distribution of pixels or noise into a coherent image. Initially, the model starts with an image consisting of pure noise and then iteratively refines this image through a series of steps, each guided by a neural network. During these steps, the model learns to remove the noise and introduce structure and details that align with the training data it has been exposed to.¹ DALL-E,² Stable Diffusion,³

0272-1716 © 2024 IEEE

Digital Object Identifier 10.1109/MCG.2024.3356169

Date of publication 19 January 2024; date of current version 25 March 2024.

and Midjourney,⁴ exemplify the impressive capabilities of diffusion models by producing a wide range of intricate images from simple text or image prompts, showcasing their ability to interpret and reproduce context, style, and subject matter.

The potential uses of AI art generators span across numerous industries and disciplines, including graphic design, advertising, architecture, fashion, and entertainment.⁵ These tools can serve as invaluable assets for designers and artists, providing them with novel ideas and creative inspiration, as well as for nondesigners who may lack the technical skills or artistic background but still wish to generate visually appealing content. Moreover, AI art generators hold the potential to democratize the creative process, making it more accessible to a wider audience, and fostering new forms of collaboration between humans and machines in the realm of artistic expression. Image synthesis models have the potential to revolutionize the field of urban design and architecture by providing instantaneous, low-cost, creative images, and site designs.

This article aims to critically evaluate the potential of AI-generated designs for common urban scenes. By developing descriptions of common urban areas with specific design elements, we aim to evaluate three popular diffusion models for this purpose. This article is structured as follows: First, we provide an overview of the current state of the academic literature surrounding AI image synthesis and tools and technologies used in urban design. Next, we outline the methodology used to evaluate AI-generated designs. Subsequently, we present and discuss our findings, focusing on the intelligibility and quality of AI-based designs. Finally, we conclude by summarizing the key insights gained from this study and discussing the potential future research directions in the field of AI and urban design.

LITERATURE FINDINGS

AI art generation has become an increasingly popular area of exploration, fueled by advancements in artificial intelligence and deep learning. Techniques such as GANs and diffusion models have been instrumental in the development and evolution of AI-generated art, pushing the boundaries of creativity and transforming how we perceive artistic expression.^{6,7}

A variety of tools and platforms has emerged to facilitate AI art generation. One such example is DALL-E, an AI-driven tool that generates images based on text inputs, showcasing the potential of AI to create visually striking and novel content.² These platforms not only democratize the art creation process but also open new avenues for exploration and

experimentation.⁸ Image synthesis tools, such as DALL-E and Stable Diffusion, are likely to become useful in various fields, including video generation and animation.⁹ Researchers are also beginning to test the capabilities of models in fields unrelated to art, such as radiology and medical imaging.¹⁰

The field of architecture commonly employs digital tools in the design process for new buildings, both for ideation and feasibility.^{11,12} The field of urban design, while classically employing freehand drawing, has also adapted to digital tools, such as ArcGIS.^{13,14} These tools provide efficient, accurate, and easily distributed models for urban planning, and can be used in many ways.^{15,16} For instance, visualization techniques, and advancing technologies such as smartphone apps, are used as methods to enhance public participation in the design process of new spaces.^{17,18} The field of urban design can benefit from the use of AI-based image synthesis models, both as ideation and iteration tools, but also as a method of democratizing the design process.

Attempts at using image synthesis models in the context of urban design are few; however, there have been notable additions to the field. Steinfeld's GAN Loci discussed the images synthesis capabilities of GANs in producing urban imagery situated in both the United States and United Kingdom.¹⁹ Recently, Senviratne et al. discussed the potential for using DALL-E to reproduce imagery scored based on style, scene, and location.²⁰ Ploennigs and Berger reviewed the body of publicly accessible prompts within Midjourney, highlighting that over 10% of all queries are related to architecture.²¹ Matias del Campo has begun to discuss theoretical work around neural art,²² while Wanyu He's XKool Technology has highlighted potential benefits of AI in design.²³

We would like to note that while the corpus of academic literature surrounding AI image synthesis is small, there have been a few foundational pieces published, and we believe much more to come as the use of these tools expands.

METHODOLOGY

Urban Design Projects

To analyze the potential abilities of AI image generation for urban design, we felt it was important to simulate specific prompting instances to test how well the AI could perform in recreating common urban planning scenes and standards. To achieve this, we developed five commonly occurring areas within a city. Within each description, we included logical named design elements that would need to be represented within the image. These items would be commonly

found in such areas and made overall sense to develop a representation of that scene (e.g., stores with signage in a commercial district).

When developing the five urban scenes, we focused on descriptiveness, and a common-sense logic for each element. Each scene was larger in scale (i.e., neighborhood as opposed to a single building) and had enough content for us to use as a reference point when scoring the corresponding AI-generated image.

AI Art Generators

DALL-E, Stable Diffusion, and Midjourney are three of the better-known tools for generating AI art. DALL-E, created by OpenAI in 2021 with a public beta release in July 2022 is a deep learning model whose release marked a turning point in the popularity and accessibility of AI generated art.²⁴ DALL-E 2 is a highly intelligent diffusion model boasting the ability to generate artwork in multiple styles, manipulate and rearrange photographs, and expand images beyond their original borders. Stable Diffusion released shortly after DALL-E was made public in 2022 and boasts similar capabilities such as stylized text to image generation and image modification.³ Midjourney, introduced in 2022 by David Holz, provides a similar output and has released five model versions since open beta.⁴ Technical details surrounding models training datasets and processes are not generally publicly available. Stable Diffusion's efforts to be transparent about the imagery used to train the model involved releasing the full dataset of imagery; however, this was not easily accessible due to size and format of the data and was removed shortly after publication.²⁵ This effort has not been replicated by either OpenAI or Midjourney. The notable differences in models lie more generally in style and features. Each model is accessed separately, Midjourney through Discord,⁴ and others on their respective websites. Dall-E is known to utilize another OpenAI product, CLIP, which links textual semantics to a visual representation.²⁶

Diffusion models, such as the ones named above, are a class of generative models that create high-quality samples, in this case images, through a process of adding and then iteratively removing noise.¹ Initially, a data sample is gradually noised over several steps until it is transformed into a state of pure noise. The model then learns a reverse process to denoise this data, effectively generating new samples from random noise.³ Transformers, an integral component in text-to-image models, process and understand textual inputs. Their architecture, based on attention mechanisms, allows for efficient handling of sequential data



FIGURE 1. Image set with test prompt generated by Craiyon (left) compared in style to an image set from Stable Diffusion.

without the need for processing in a sequential order.²⁷ In this context, transformers encode textual information into a high-dimensional space, which guides the diffusion process to produce images that align with the textual description.²⁷ This integration of transformers and diffusion models facilitates the generation of relevant imagery. More simply, transformers used in image generation facilitate the model's ability to capture context and nuance of the text input to reach the desired image output.

Aside from these options, we tested a variety of tools with a separate generic prompt (e.g., instructing the generators to create a mockup of a "healthy, walkable neighborhood design" to test the feasibility of this research). We looked at programs that were popular and well recognized in the AI community, noting a range of parameters (e.g., cost, ability, and consistency). We considered DALL-E 2, Stable Diffusion, and Midjourney as being cost effective, having an established community, and able to produce images in the level of detail, realism, and quality that we desired. Art generators that could produce realistic imagery were an intentional choice for our research; however, this is not an absolute requirement outside of this work. Models that produced art in a singular, specific style, such as Craiyon, were not selected for our study due to lack of realism²⁸ (see Figure 1).

Study and Evaluation Process

Run Structuring

We conducted four generations of each of the five paragraph prompts through both DALL-E 2, Stable Diffusion, and Midjourney, which we called a run. Each model produced four images per prompted run resulting in 16 AI generated images per paragraph prompt, per AI generator, for a total of 240 images. We chose to complete multiple runs with identical prompts instead of single runs to eliminate the possibility of

any epistemic luck on behalf of the AI, as well as to spot areas of consistency in the generator’s interpretation of the prompt and quality of output. These runs were completed back-to-back in rapid succession and each image was assigned a number corresponding to its human designed project and run number. Any individualized parameter or setting on the models were left as default to not give advantage to a particular model based on features. Each model was only prompted with the exact text provided in Table 1, and the first 16 images the model provided for each prompt were downloaded to be scored.

Image Scoring

Computational quantitative assessment of text-to-image generation has been demonstrated²⁹; however, this methodology ignores aesthetic appeal, human preference, and other important qualitative metrics, especially for non-AI trained urban designers.²⁰ Researchers have been challenged to assess AI-generated images for human preference without introducing bias, subjectivity, and oftentimes a lack of replicability. Urban design scoring systems, such as the LEED v4 for Neighborhood Development certification, are widely used in the industry, but require quantitative metrics associated with specific projects, such as coverage percentages, sidewalk widths, and/or building heights. This makes existing methods difficult to use effectively for evaluating images without underlying numeric details.

The need for a more open-ended evaluation process capable of measuring imaged-based AI outputs led us to adopt a method created by Google in 2020—the sensibleness and specificity average (SSA). Google developed the SSA to assist in evaluating the quality of Meena, a conversational AI chatbot and structured the method as a simple and flexible way to evaluate the success of AI in mimicking desired human behavior.³⁰ We adapted the two core components of this metric to accurately capture both the overall sensibleness of an AI-generated image, as well as the specific design elements that would need to be reflected in each image. The SSA’s effectiveness in our context lies in its dual focus on general coherence and detailed accuracy. Sensibleness, in the realm of image-based AI for architecture and urban design, pertains to the overall realism and feasibility of the generated images. It ensures that the AI does not produce fantastical or implausible designs, maintaining a level of practicality essential in urban planning. Specificity, on the other hand, assesses the extent to which the images incorporate specific, requested design elements, reflecting the AI’s ability to adhere to detailed

TABLE 1. Urban scene seed text.

Scene Name	Seed Text
Commercial District	A bustling commercial shopping district during daytime, showcasing a variety of stores with vibrant signage , pedestrians, and street vendors. The architecture includes a mix of modern and traditional styles, with wide, clean sidewalks, a public transit stop, and bike lanes .
Industrial Area	An expansive industrial area, characterized by large, metal warehouses, smokestacks emitting steam , a network of railway tracks for cargo trains, and cranes lifting containers. The area should include concrete and fencing around a perimeter .
Office Complex	A modern mixed-use office complex, featuring a tall office building with glass facades , a landscaped plaza with benches and modern art sculptures , people in business attire walking and conversing, and a small outdoor café with tables and umbrellas . The complex is surrounded by green spaces and has a sleek, contemporary design.
Urban Park	A lively urban park in the afternoon, filled with diverse activities, including a children’s playground with slides and swings, a small pond with ducks , walking paths surrounded by lush greenery and flowers, and groups of people picnicking on the grass. The park is framed by tall trees and has a backdrop of city skyscrapers in the distance.
Residential Neighborhood	A tranquil residential neighborhood at sunset, featuring a row of houses with front porches, well-manicured lawns with colorful flowerbeds, a winding sidewalk lined with streetlamps , and parked cars of various models.

project requirements and constraints. This also allows us to highlight where AI models hallucinate, or reliably produce design elements. By adapting the SSA, we can holistically evaluate these generative models, ensuring they not only produce realistic and coherent images but also respect the intricacies and specificity inherent in architectural and urban design.

Each AI-generated image was rated on a scale of zero to two for sensibleness. The scale ranged from a zero, an image that was not logical, had difficult to decipher components, or did not clearly communicate a scene, to a two, an image that had few or no illogical



FIGURE 2. Stable Diffusion generated images by sensibleness score from left to right: score of 0 (top), score of 1 (top), score of 2 (bottom).

components, made overall sense, and could clearly represent the intended scene. We assigned a score of one if an image was a mixture of the two extremes (see Figure 2).

Each scene was also rated for specificity on a scale of 1–4, with each point being awarded for a specific component of the paragraph prompt it was given (see bold terms in Table 1). For example, when evaluating the specificity of images generated for the commercial district, we were looking for instances of stores with signage, sidewalks, public transit stops, and bike lanes. While we did not expect the model to perfectly capture these concepts in evaluation, we were looking for arguable attempts at these elements (see Figure 3).

Evaluation Process

Two evaluators participated in a blind review of each image and recorded their perceived scores. Both reviewers are current Ph.D. students within urban planning and design with relevant experience using generative AI tools. The reviews were completed separately from one another to reduce any risk of confirmation bias or influence in scoring by the other reviewer. Each evaluator was familiarized with the SSA, related urban design project examples, and goals of the study before participation. The scoring process happened over the course of two weeks, including training and blind review. Images were scored in a



FIGURE 3. DALL-E Images for industrial area prompt with a specificity score of 1 (left) and 2 (right), top row and 3 (left), and 4 (right) bottom row.

randomized fashion, with the necessary components for specificity scoring included side by side with each image.

ANALYSIS

For our main analysis, scores for each image were averaged between the two evaluators. We performed a secondary, independent analysis of reviewers, highlighted in the end of the results. To compute descriptive statistics for our interval variables, we calculated means and standard deviations. Due to the relatively small sample size ($N = 240$), determining the distribution of both sensible and specific was important to choose an appropriate statistical method for comparison. We performed Shapiro–Wilk tests, which showed the data departed significantly from normality ($W = 0.776$ and 0.858 , $p < 0.001$) for sensible and specific, respectively. Based on this outcome, further comparisons were performed with nonparametric testing methods. We compared SSA metrics across models and projects using independent sample Kruskal–Wallis tests.

RESULTS

Descriptives

In total we scored 240 images, split evenly between three art generators ($n = 80$ for each). When analyzing the three art generators together, the mean score for

TABLE 2. Descriptives for AI art generators and projects.

AI Art Generator	Mean Score Sensible	Mean Score Specific
DALL-E 2	1.044 (0.612)	2.625 (0.891)
Stable Diffusion	1.238 (0.443)	2.500 (0.694)
Midjourney	1.556 (0.477)	2.888 (0.711)
Commercial District	1.198 (0.481)	2.291 (0.743)
Industrial Area	1.323 (0.455)	2.708 (0.771)
Office Complex	1.469 (0.559)	2.583 (0.679)
Urban Park	1.115 (0.638)	3.083 (0.846)
Residential	1.292 (0.582)	2.688 (0.689)
Combined	1.279 (0.556)	2.671 (0.784)

sensible was 1.279 (SD = 0.556) and the mean score for specific was 2.671 (SD = 0.784). When split by generator, all three models performed similarly (see Table 2). Also included are mean scores split by urban design project.

Generator and Project Comparisons

A Kruskal–Wallis test showed that there was a significant difference of means for both sensible and specific averages ($H = 33.873$, $p < 0.001$ and 10.455 , $p = 0.005$). We then conducted post hoc tests to test pairwise comparisons. We found that Midjourney scored significantly higher as compared to both Stable Diffusion ($p < 0.001$) and DALL-E ($p < 0.001$). Stable Diffusion also scored significantly higher as compared to DALL-E ($p = 0.048$).

A Kruskal–Wallis test was performed to evaluate whether the SSA metrics differed by project. The results indicated a significant difference for both metrics. On the metric of sensible, there was a significant difference between groups ($H = 10.100$, $p = 0.039$). Post-hoc testing was performed to evaluate the differences between each project. The mean sensible score for most projects showed nonsignificant differences (p 's > 0.112). However, the office complex project had a significantly higher sensibleness score as compared to the urban park ($p = 0.015$). On the metric of specific, there was another significant difference between groups ($H = 24.122$, $p < 0.001$). Post-hoc testing was performed to evaluate the differences between each project. Here, the urban park scored significantly higher than both the commercial district ($p < 0.001$) and the office complex ($p = 0.011$).

Design Element Analysis

To evaluate differences in overall model ability to produce specific design elements within a prompt, we scored design elements individually within each

TABLE 3. Descriptives for design elements.

Urban Scene	Design Element	Percent Appearance
Office Complex	Tall glass buildings	100.00%
Residential Neighborhood	Lawns with flowers	100.00%
Residential Neighborhood	Row of houses	97.92%
Urban Park	Tall trees	97.92%
Commercial District	Stores with signage	97.92%
Commercial District	Bike lanes	33.33%
Industrial Area	Fencing	33.33%
Residential Neighborhood	Parked cars	33.33%
Commercial District	Public transit stop	14.58%
Office Complex	Modern art sculptures	10.42%

image. This allowed us to track whether particular elements appeared more or less frequently within the overall image set. The top and bottom appearing five elements are shown in Table 3.

DISCUSSION

Overview of Results

Our study assessed the performance of three AI art generators—DALL-E 2, Stable Diffusion, and Midjourney—in creating images for urban design projects. We evaluated these images based on two metrics: sensibleness and specificity, applying these criteria across different urban design scenes. Overall, our findings reveal significant variations in the performance of these generators, both in general and in relation to specific project types.

Generator Performance

The mean scores for sensible and specific qualities varied significantly among the three AI art generators. Midjourney outperformed the other two generators in both metrics, with particularly high scores in the specificity category. Stable Diffusion, although scoring lower than Midjourney, was significantly more effective than DALL-E 2. These results were statistically supported by a Kruskal–Wallis test and subsequent post-hoc tests. This suggests that while all generators are capable of creating relevant urban design imagery, certain generators are more adept at generating images that align closely with specific design elements. The higher performance of Midjourney, especially, indicates its superior ability to interpret and

visualize detailed and specific aspects of urban design prompts.

Project-Specific Analysis

The performance of the AI art generators also varied significantly across different urban design projects. Interestingly, the office complex project stood out with a significantly higher sensibleness score, particularly when compared to the urban park project. Conversely, the urban park project scored highest in specificity, indicating its strong alignment with specific design elements. This discrepancy could be due to the inherent characteristics of the different urban environments. For instance, office complexes might have more standardized and recognizable features, making them easier to render sensibly. On the other hand, urban parks, with their diverse natural elements and layouts, might offer more scope for specificity in design. This may also speak to the training data that these models are based on. There is a clear bias toward specific types of architecture throughout the imagery, especially western design.

Design Element Analysis

Our analysis of individual design elements within each urban scene demonstrates the relationship between training data and novel output. Certain elements like tall glass buildings in office complexes and lawns with flowers in residential neighborhoods appeared consistently across images. This would likely be explained by the representation of these elements within the training data for the models. Here the model would use learned probabilities to progressively transform a noisy sample into a coherent image that aligns with its training on tall glass buildings. Elements such as modern art sculptures in office complexes and public transit stops in commercial districts appeared less frequently. This suggests a potential gap in the training data used for these models, and therefore the probability of accurately including specific or less common elements in urban designs.

Visualizing Urban Space

We believe the diffusion model, and AI art as a whole, is a powerful potential tool to visualize urban space. Imagery could be used in prototyping or initial design of a new space, to engage the public more effectively, to simulate changes in urban form due to development, or to integrate a particular historic or cultural architectural style within an existing setting. Tools such as Dall-E, Stable Diffusion, and Midjourney are publicly available, low cost, and rapid. It follows that

these tools will become more frequently used within the design and planning industries. Therefore, it is crucial to critically assess the outputs of the models before they become standard practice. The use of the SSA metric has several strengths, as it relates to urban planning imagery. First, the SSA can facilitate constructive, iterative feedback. The metric provides a focused, easy to understand way to provide criticism of a complex model. If an image lacks sensibleness, it indicates the need for better grounding in realistic urban principles. If it lacks specificity, this suggests the AI model needs refinement or a more diversified training dataset. The SSA can also provide a framework for urban designers to interface with the public. Sensibleness ensures imagery is understandable and practical from a nontechnical audience's viewpoint. Specificity would ensure more detailed requirements are visually represented. Finally, the SSA could also serve as a benchmarking tool to compare AI imagery against existing urban designs. Sensibleness provides a metric to measure whether AI creativity overrides existing urban planning principles, while specificity could be used to ensure that AI design does not become overly generic, instead tailored to a project's unique requirements.

Limitations and Future Research

Despite the valuable insights gained from this research, there are several limitations that need to be acknowledged. First, the sample size was relatively small, with 240 images analyzed across three AI art generators and five urban scenes. This may limit the generalizability of the findings, as a larger sample size could potentially reveal different patterns or outcomes in the data. Second, the study only focused on realistic AI art, which does not fully represent the range of model capabilities. Including more prompts specifying style (i.e., mock-up, site drawing, site layouts, etc.) in future research could help to draw more robust conclusions about the performance of AI-generated art in the context of urban design. Third, the study relied on human evaluators to assess the sensibleness and specific metrics, which introduces a level of subjectivity to the results. Although the evaluators were found to be consistent in their assessments, different evaluators might have different interpretations or preferences, which could influence the results.

AI image generation inherently carries the risk of perpetuating and amplifying existing biases present in training datasets. Models learn to generate images based on the data they are fed, which often reflects historical, cultural, and socio-economic biases. If a

model is predominantly trained on images of urban landscapes from affluent, Western cities, it may disproportionately generate designs that echo these contexts, neglecting the diversity of architectural styles and urban planning principles found globally. This can lead to a homogenization of design ideas, overlooking the unique needs and aesthetics of different regions, cultures, and socio-economic groups. Additionally, such models might underrepresent or misrepresent non-Western, indigenous, or less economically developed urban environments, potentially reinforcing stereotypes or contributing to a cultural erasure in design. This bias not only limits the creativity and applicability of the generated designs, but also raises ethical concerns about inclusivity and representation in the field of architecture and urban planning.

Further analysis could incorporate various other metrics to assess the quality of the images, rather than relying on only the SSA. This could include design bias metrics. A secondary comparison to evaluate and compare AI-generated and human-generated designs could provide a deeper understanding of discrepancies. By understanding the limitations in depicting certain design elements, we can guide the development of more sophisticated AI models, tailored to the nuanced demands of urban design. This can eventually lead to more efficient and accurate tools for architects, urban planners, and designers, enhancing the planning and visualization process in urban development projects.

ACKNOWLEDGMENTS

The authors would like to acknowledge the use of DALL-E 2, Stable Diffusion, and Midjourney in creating all imagery for this research. This work was supported in part by the National Science Foundation (#2125858), in part by USDOT Consortium of Cooperative Mobility for Competitive Megaregions, and in part by UT Austin Good Systems, and the MITRE Corporation.

REFERENCES

1. P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," *Adv. Neural Inf. Process. Syst.*, vol. 34, pp. 8780–8794, 2021.
2. A. Ramesh et al., "Zero-shot text-to-image generation," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 8821–8831.
3. R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 10674–10685.
4. Midjourney Inc., "Midjourney documentation." 2023. [Online]. Available: <https://docs.midjourney.com>
5. J. Jiao and J. Mallot, "AI image generation for architecture." 2022. [Online]. Available: <https://smartcity.tacc.utexas.edu/design>
6. J. Agnese, J. Herrera, H. Tao, and X. Zhu, "A survey and taxonomy of adversarial neural networks for text-to-image synthesis," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 10, no. 4, 2020, Art. no. e1345.
7. Z. Chen, L. Chen, Z. Zhao, and Y. Wang, "AI illustrator: Art illustration generation based on generative adversarial network," in *Proc. IEEE 5th Int. Conf. Image, Vis. Comput.*, 2020, pp. 155–159, doi: [10.1109/ICIVC50857.2020.9177494](https://doi.org/10.1109/ICIVC50857.2020.9177494).
8. J. Brusseau, "Acceleration AI ethics, the debate between innovation and safety, and stability AI's diffusion versus OpenAI's Dall-E," 2022, *arXiv:2212.01834*.
9. L. Wang, W. Chen, W. Yang, F. Bi, and F. R. Yu, "A state-of-the-art review on image synthesis with generative adversarial networks," *IEEE Access*, vol. 8, pp. 63514–63537, 2020.
10. L. C. Adams, F. Busch, D. Truhn, M. R. Makowski, H. J. W. L. Aerts, and K. K. Bressen, "What does DALL-E 2 know about radiology?," *J. Med. Int. Res.*, vol. 25, 2023, Art. no. e43110, doi: [10.2196/43110](https://doi.org/10.2196/43110).
11. P. Szalapaj, *Contemporary Architecture and the Digital Design Process*. Evanston, IL, USA: Routledge, 2014.
12. J. Walliss and H. Rahmann, *Landscape Architecture and Digital Technologies: Re-Conceptualising Design and Making*. Evanston, IL, USA: Routledge, 2016.
13. T. Bradecki and M. Stangel, "Freehand drawing for understanding and imagining urban space in design education," *Architecture Civil Eng. Environ.*, vol. 7, no. 2, pp. 5–14, 2014.
14. N. Chigbu and M. Daberechi, "Site view reconstruction for urban planning using ArcGIS, Google sketch up and Google earth a case study of the University of Nigeria Enugu campus." 2011.
15. F. Chaaban et al., "Using ArcGIS modelbuilder and aerial photographs to measure coastline retreat and advance: North of France," *J Coast Res.*, vol. 28, no. 6, pp. 1567–1579, 2012.
16. K. Ulm and X. Wang, "Efficient reality-based 3D city modeling with cybercity-modeler-management in arcgis (ESRI) and visualization with terrainview," in *Proc. Workshop Papers*, 2005, Art. no. 57.
17. K. Al-Kodmany, "Using visualization techniques for enhancing public participation in planning and design: Process, implementation, and evaluation," *Landscape Urban Plan.*, vol. 45, no. 1, pp. 37–45, 1999.
18. A. Wilson, M. Tewdwr-Jones, and R. Comber, "Urban planning, public participation and digital technology: App development as a method of generating citizen involvement in local planning processes," *Environ. Plan B Urban Anal. City Sci.*, vol. 46, no. 2, pp. 286–302, 2019.

19. K. Steinfeld, "GAN loci," ACADIA, 2019. Accessed: Nov. 27, 2023. [Online]. Available: https://papers.cumincad.org/data/works/att/acadia19_392.pdf
20. S. Seneviratne, D. Senanayake, S. Rasnayaka, R. Vidanaarachchi, and J. Thompson, "DALLE-URBAN: Capturing the urban design expertise of large text to image transformers," in *Proc. Int. Conf. Digit. Image Comput., Techn. Appl.*, 2022, pp. 1–9, doi: [10.1109/DICTA56598.2022.10034603](https://doi.org/10.1109/DICTA56598.2022.10034603).
21. J. Ploennigs and M. Berger, "AI art in architecture," *AI Civil Eng.*, vol. 2, Aug. 2023, Art. no. 8, doi: [10.1007/s43503-023-00018-y](https://doi.org/10.1007/s43503-023-00018-y).
22. M. Del Campo, "This City does not exist an attempt at a theory of neural urban design," *FORUM A+P Interdiscipl. J. Architecture Built Environ.*, no. 23, pp. 36–40, Oct. 2021, doi: [10.37199/F40002303](https://doi.org/10.37199/F40002303).
23. W. He, "Urban experiment: Taking off on the wind of AI," *Architectural Des.*, vol. 90, no. 3, pp. 94–99, May 2020, doi: [10.1002/AD.2574](https://doi.org/10.1002/AD.2574).
24. W. Knight, "Where the AI art boom came from and where it's going." 2023. [Online]. Available: <https://www.wired.com/gallery/where-the-ai-art-boom-came-from-and-where-its-going/>
25. A. Baio, "Exploring 12 million of the 2.3 billion images used to train stable diffusion's image generator," Waxy. Accessed: Jan. 03, 2024. [Online]. Available: <https://waxy.org/2022/08/exploring-12-million-of-the-images-used-to-train-stable-diffusions-image-generator/>
26. A. Radford et al., "Learning transferable visual models from natural language supervision," Feb. 2021. [Online]. Available: <http://arxiv.org/abs/2103.00020>
27. A. Vaswani et al., "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
28. CraiyonLLC, "Craiyon, AI image generator." 2023. [Online]. Available: <https://www.craiyon.com>
29. N. Gu and P. Amini Behbahani, "A critical review of computational creativity in built environment design," *Buildings*, vol. 11, no. 1, Jan. 2021, Art. no. 29, doi: [10.3390/BUILDINGS11010029](https://doi.org/10.3390/BUILDINGS11010029).
30. D. Adiwardana et al., "Towards a human-like open-domain chatbot," 2020, *arXiv:2001.09977*.

CONNOR PHILLIPS is currently working toward the Ph.D. degree in city and regional planning with the University of Texas at Austin, TX, 78712-1710, USA. His research in the Urban Information Lab at UT includes generative AI, ethics, and smart cities. He is the corresponding author of this article. Contact him at connorphillips@utexas.edu.

JUNFENG JIAO is an associate professor with the Community and Regional Planning Program, University of Texas at Austin, TX 78712-1710, USA. His research focuses on smart cities, smart transportation, urban informatics, and ethical AI. Contact him at jjiao@austin.utexas.edu.

EMMALEE CLUBB received the master's degree in information systems from the University of Texas at Austin, TX 78712-1710, USA. Her research interests are focused on the intersection of ethics, empathy, and human design within technology. Contact her at emmaleeclubb@utexas.edu.