AI-aided Design? Text-to-image Processes for Architectural Design

Matteo Flavio Mancini, Sofia Menconero

Abstract

Artificial Intelligence (AI) is marking a turning point in many aspects of human life, and it is appropriate to question its potential use in the architectural representation processes.
This contribution provides a brief overview of the recent past of AI technologies to explain how they work, a snapshot of the current state of the art from text-to-image processes to image-to-3D processes, mainly focusing on the StableDiffusion *platform.*
It also offers an overview of the latest studies in the field of architectural design. The subsequent experimentation becomes an opportunity to showcase the potential of AI in the co-creation process and the ability to simulate various graphic techniques, up to photorealistic visualization. On the other hand, it presents the limitations that, at the current stage of development, sometimes invalidate the results of text-to-image processes concerning the scientific aspects of representation.
The conclusions reflect on the differences between human and artificial intelligence, the theme of shared authorship between humans and machines, and their consequences for architectural design.

Keywords: artificial intelligence, text-to-image, design drawing, authorship, stablediffusion.

Introduction

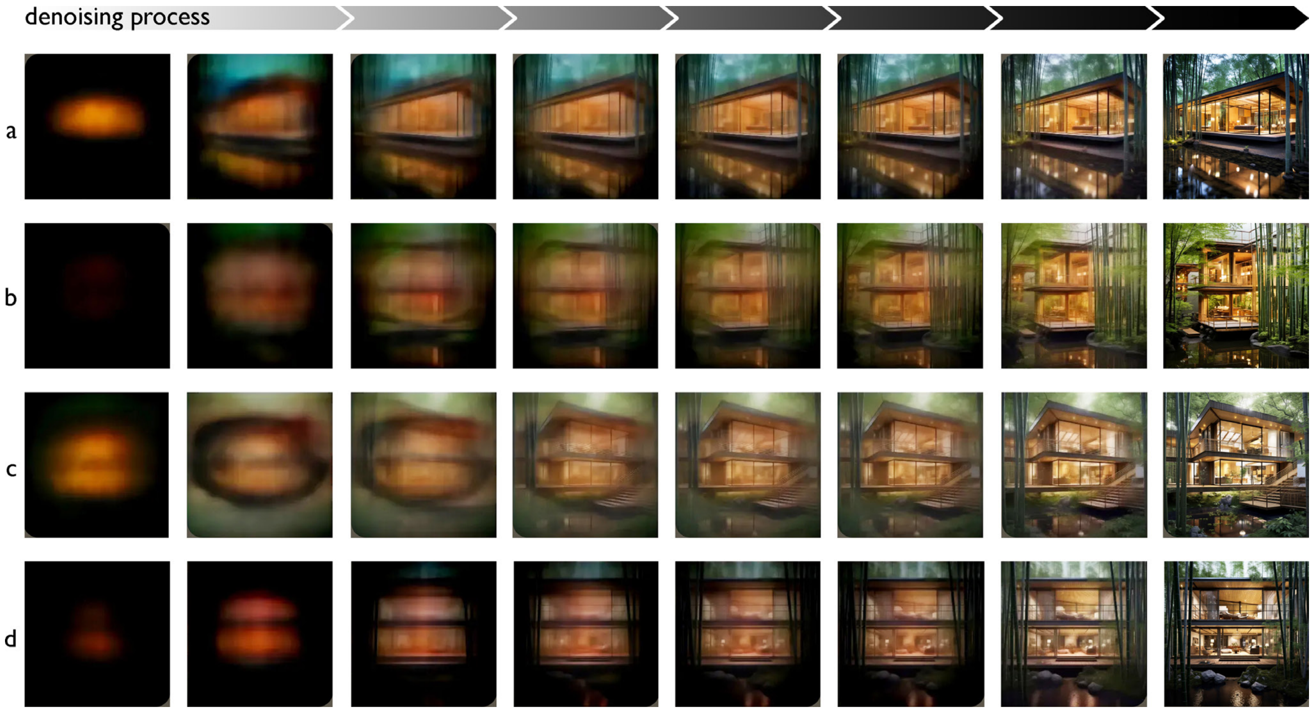
Architecture and architectural drawing have undergone significant developments over the past thirty years. The *first digital turn* [Carpo 2013] introduced digital representation in the 1990s, while the *second digital turn* [Carpo 2017] began with the spread of algorithms and big data in the 2010s. Ten years later, we are witnessing another potential turning point due to the sudden development and diffusion of Artificial Intelligence (AI) tools, which are already in use in major architectural firms such as Coop Himmelb(l)au [Prix et al. 2022], Zaha Hadid Architects [Wallish 2022], and Foster + Partners [Tsigkari et al. 2021].
One branch of AI, based on text-to-image processes, offers easily accessible solutions dedicated to image creation. This machine-learning model uses descriptive natural language as input and produces an image based on the provided description. The results obtained from these platforms are remarkable in matching the entered textual prompts and the flexibility of graphic techniques they can (re)produce.
Starting from the assumption that, at the current state, these AIs have no creative consciousness or actual ability to understand compositional and projective rules or the spatiality represented in the images, it is still worth exploring their potential use in architectural representation processes. With this goal in mind and considering the intrinsic characteristics of this technology, which will be explained in the follow-

ing paragraphs, the experimentation of text-to-image AI through the open-source platform *StableDiffusion* is proposed to create perspective images capable of contributing to the preliminary stages of project conception.

(Recent) past and present of AI-based image generation

Generative AI systems in the field of architecture and design are rapidly advancing. A pivotal moment in image generation research was marked by the invention of Generative Adversarial Networks (GANs) in 2014 [Goodfellow et al. 2014]. This represents a deep learning architecture where two opposing neural networks, a generator and a discriminator, interact iteratively during training to reach a point where the discriminator can no longer differentiate between synthetic images produced by the generator and real images used as training data. In 2016, a GAN architecture capable of generating plausible images from detailed textual descriptions was developed [Reed et al. 2016], effectively initiating the AI text-to-image system. Another significant advancement came with a more efficient natural language processing-based learning method known as CLIP (Contrastive Language-Image Pre-training) [Radford et al. 2021]. This image classification model identifies objects by learning from text associated with an image, rather than relying on manually assigned labels, and it was trained on 400 million image-text pairs extracted from the web. CLIP models can estimate the conformity of a generated image to a textual prompt [Colton et al. 2021].

Fig. 1. Denoising process during the generation of four variants (a, b, c, d) in Midjourney *through the prompt: a modern Japanese house in a bamboo forest in spring (authors' processing).*



An example of this pairing is the image generation system called *VQGAN-CLIP*, which utilizes an even more powerful GAN neural network. The significant contributions of the *VQGAN-CLIP* architecture include visual quality in both image generation and manipulation, semantic fidelity between text and the generated image, efficiency due to not requiring additional training beyond pre-trained models, and the value of open development and scientific progress [Crowson et al. 2022, p. 2]. Subsequently, GAN-based systems were replaced by diffusion models, which are probabilistic machine learning models trained to remove noise previously introduced from images by learning to reverse the diffusion process [Dhariwal et al. 2021]. The training of these models enables them to utilize denoising methods to synthesize new noise-free images from random inputs (fig. 1).

Some applications of AI in the field of architecture and design include:

- Text-to-image: the most common operation involves generating images based on textual descriptions, often combined with other functionalities.
- Image-to-image: transforming an input image to match the characteristics of a target image. This can be used for style transfer, object manipulation (inpainting), converting black and white images to color, or increasing image resolution (upscaling).
- Text or image-to-video: creating videos from textual prompts (e.g., *Make-a-Video* [Singer et al. 2022] or *CogVideo* [Hong et al. 2022]) or generating animations by editing images generated through image-to-image (e.g., *Deforum*), producing effects similar to stop-motion videos.
- Text or image-to-3D: generating 3D models from textual prompts (e.g., *Point-E* for point clouds [Nichol et al. 2022] or *Shape-E* for textured meshes [Jun et al. 2023]), or generating 3D models from images (e.g., *Kaedim*).

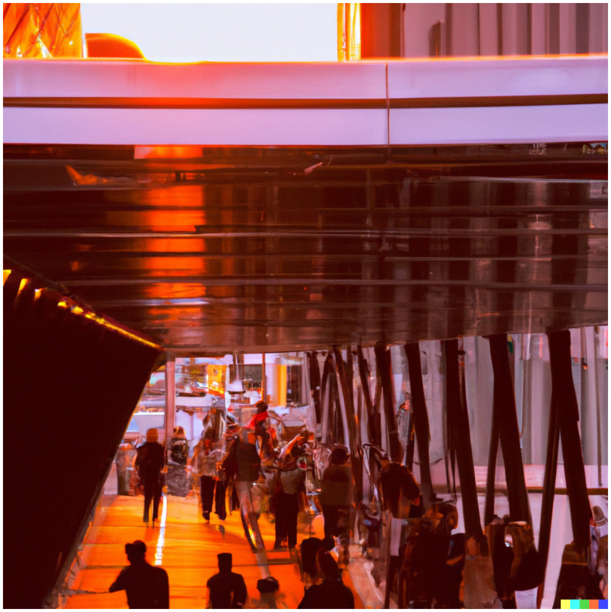
The recent incredible proliferation of AI text-to-image capabilities can be attributed to the activation of user-friendly platforms, even for non-expert users, such as *DALL-E 2*, *Midjourney*, and *StableDiffusion*.

The main platforms for text-to-image AI

DALL-E is the first of the three platforms introduced in January 2021 (the current version 2 was released in April 2022) by *OpenAI* [Ramesh et al. 2021], the same developers of *ChatGPT*. The platform, available for online subscription, offers four functions: the generation of realistic and artistic images from a textual description that can combine concepts, attributes, and styles (fig. 2); outpainting, which involves extending the image beyond its original boundaries by creating a new composition; inpainting, which allows the modification of image portions by adding or removing objects through a textual description while maintaining consistency with the rest of the scene; and the generation of variations inspired by an input image.

Midjourney, released on July 12, 2022, has now reached version 5.2 with significant improvements in terms of adherence to prompts and photorealism compared to its initial release (fig. 3). After a year, it boasts over 15 million users [1]. Like *DALL-E*, it is available for online subscription. The three main generative activities on *Midjourney* are: generating an image from a textual prompt, generating a description from an image, and

Fig. 2. Image generated with DALL-E 2 *through the prompt: a modern building on a crowded street at sunset (authors' processing).*



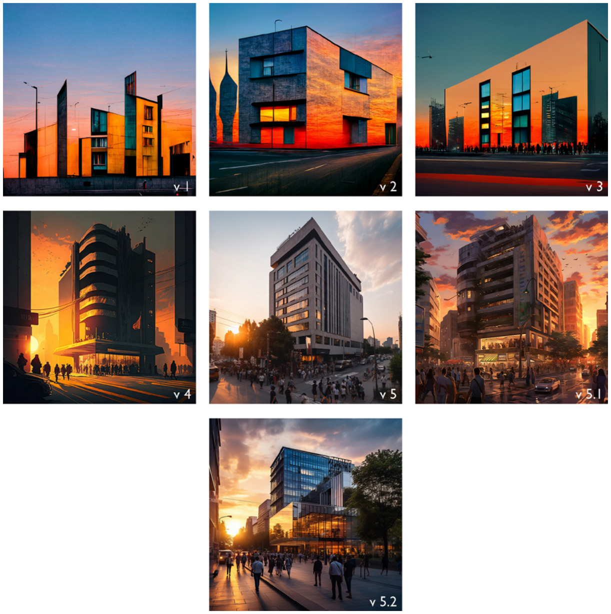


Fig. 3. Comparison between different images generated from the same text (prompt: a modern building on a crowded street at sunset) for different versions of Midjourney *(authors' processing).*

Fig. 4. Comparison between different images generated from the same text (prompt: a modern building on a crowded street at sunset) for different versions of StableDiffusion *(authors' processing).*

synthesizing an image from two to five input images. *Midjourney* (like *StableDiffusion*) also allows the use of a negative prompt in case specific elements are not desired in the generated image.
StableDiffusion, released in August 2022, is the only one among the three platforms to be open-source and is based on a diffusion model called *latent diffusion model* [Rombach et al. 2022]. The current beta XL version is only available online through a subscription, while previous versions can be installed locally for free. *StableDiffusion* supports image generation using a text prompt describing the elements to include or exclude from the output (fig. 4), inpainting and outpainting, image-to-image generation, and upscaling. It is also possible to add extensions to *StableDiffusion*, such as *ControlNet*, which generates variations of an input image through textual descriptions, and *Deforum*, which, through the image-to-image function, generates a series of images with minor transformations and stitches them together to create a video.
The biggest advantage of *StableDiffusion* over the other platforms is that end-users can implement additional training (fine-tuning) to optimize generation outputs to specific use cases. For example, in architectural studies where AI has become part of the creative process, the neural network is trained with targeted images from the studio's design repertoire to achieve results more in line with architectural and graphic language.
Therefore, unlike the previous platforms, *StableDiffusion* allows greater freedom in terms of customization of the generative process, which is why it was chosen for further experimentation, coupled with the *ControlNet* extension [Zhang et al. 2023], which improves output control. The latter is a neural network structure designed to handle diffusion models by incorporating additional conditions: by manipulating the input conditions of the blocks, it further controls the overall

behavior of an entire neural network. *ControlNet* operates based on an input image and a textual description, allowing the generation of images that conform compositionally to the input but also follow the specified description. The data processing involves, first and foremost, the generation of a map based on the input image (called the annotation or preprocessing phase), which is used by the network to generate variants with the described textual characteristics (fig. 5).

Back and forward in latent space: between training and generation in the *StableDiffusion* model

To properly approach the experiments that will follow in the next paragraphs, it is essential to try to understand not so much the strictly technical-computer aspects but the processes carried out by this type of AI, specifically *StableDiffusion*, in the two distinct moments of training and generation, as both appropriate use and critical interpretation of this technology depend on them.
Diffusion models borrow from thermodynamics the concept of diffusion, which is the phenomenon whereby particles of a fluid move randomly within another fluid with a different concentration until they reach a new equilibrium condition. Similarly, AI-generated images during generation progressively emerge from the chaos of digital noise. The principle of diffusion is used both in the training phase (forward diffusion) and in the generation phase (reverse diffusion). In *StableDiffusion*, both processes occur in the latent space, a numerical/informational space in which images are translated into tensors (multi-dimensional matrices) to work on a compressed version of them, lighter than the initial pixel space of the images. Texts describing the images also undergo a similar translation and compression. The analogy between the latent representations of texts and images is important because it helps understand that AIs do not store and process collections of syllables, words, or portions of images but operate on abstract numerical representations of image features, represented objects, possible environmental situations, and various techniques and styles. The

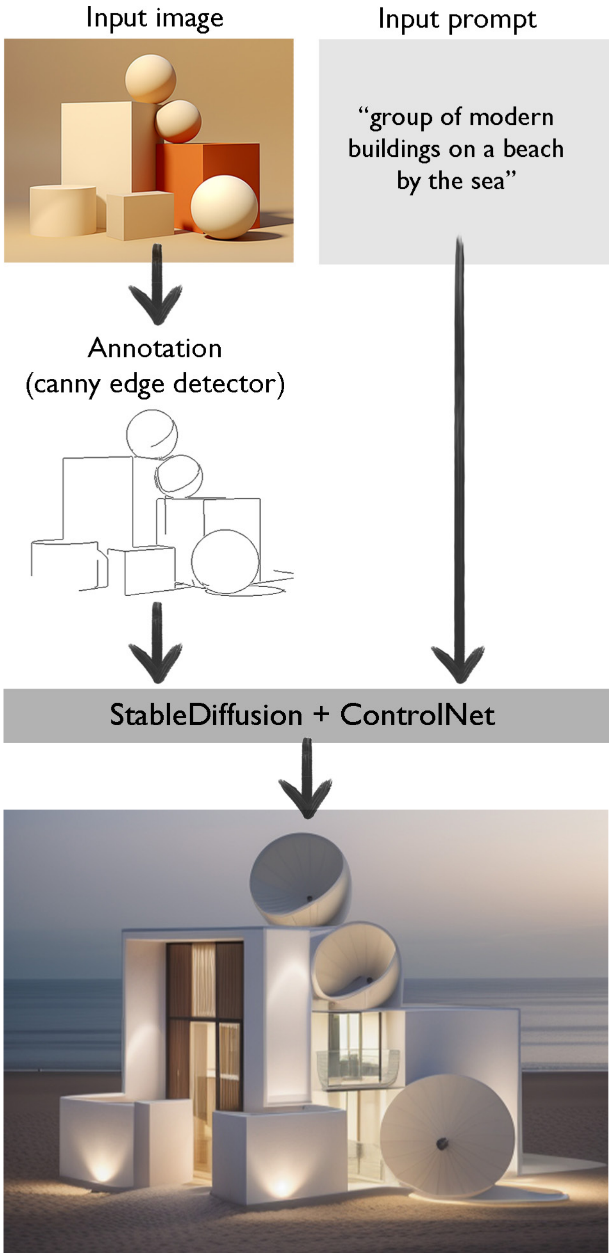
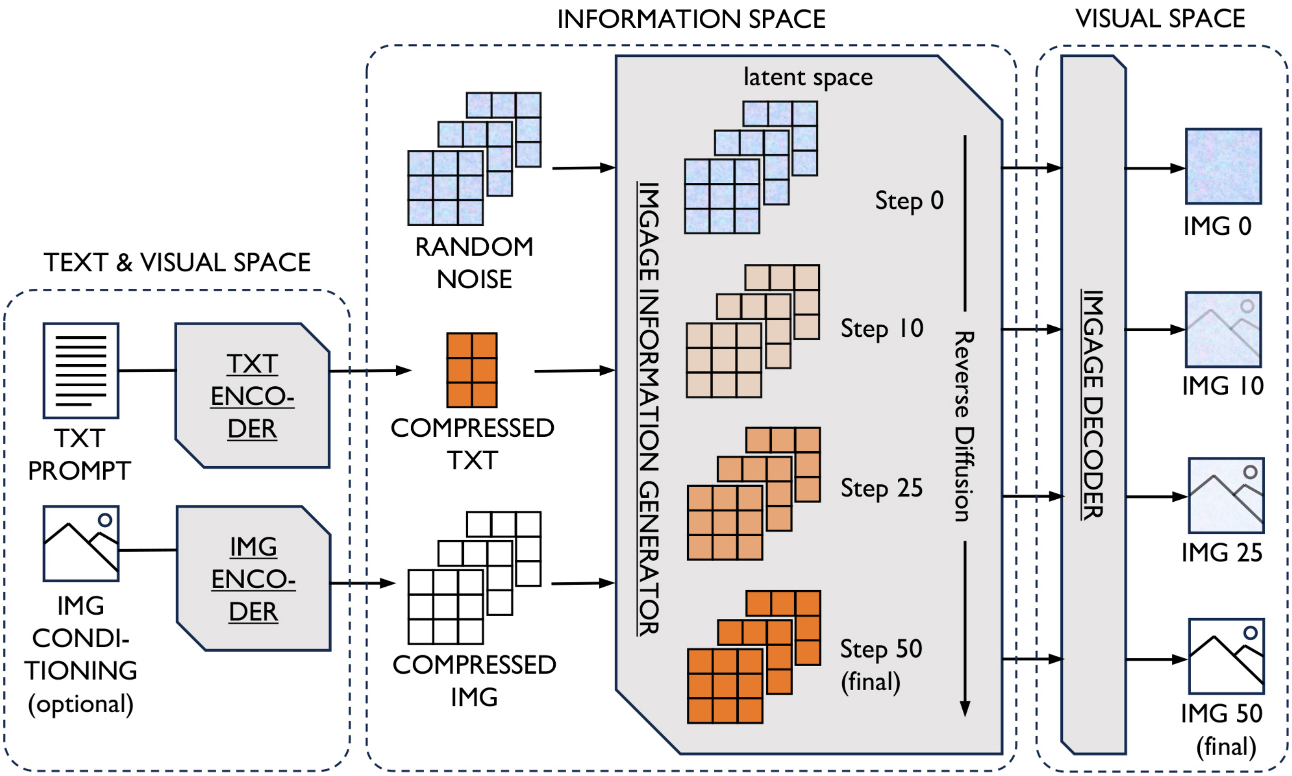


Fig. 5. Diagram of image generation through StableDiffusion *with ControlNet conditioning (canny edge) (authors' processing).*

latent space can be imagined as the place where the AI stores, during training, and retrieves, during generation, its 'knowledge.'
The training of these AIs is not progressive over time but occurs before publication, so their 'knowledge' is static and periodically updated with the progression of versions. For example, *StableDiffusion* is trained on the open-access dataset LAION 5B, composed of 5 billion image-caption pairs whose content is explorable, starting from the input of a textual prompt, through a dedicated portal [2]. Consulting the training dataset allows one to get an idea of the correspondences between terms and images and, therefore, what can be expected from the results of the processing: a search that does not return consistent results indicates that images corresponding to the expectations for that textual input cannot be generated. During training, images from the database are processed by introducing random noise patterns of different intensities. The images thus processed, together with the corresponding captions, are subjected to the AI to train it to identify the type of pattern adopted, the amount of noise introduced and to remove both to improve the quality of the images. In this way, through forward diffusion, the AI simultaneously learns how to obtain noise-free images and the correspondences between images and texts.
What is learned is used by *StableDiffusion* to develop a generative process that starts and ends in a space where data (texts and images) are suitable for human

Fig. 6. Diagram of the image generation process through the diffusion model adopted by StableDiffusion *(authors' processing).*



perception, passes through a purely informational space (latent space) where data are represented by tokens (texts) and tensors (images).
The generative process can be divided into three fundamental blocks (fig. 6). The first involves the compression and transformation into numerical information through encoders (specifically trained neural networks) of the input data entered to condition image generation. In *StableDiffusion*, thanks to the *ControlNet* extension, textual inputs can be supplemented with optional graphical conditions. In the second block, through the reverse diffusion process, the processing of the inputs in relation to known knowledge takes place. This iterative process goes through several denoising steps to refine the correspondence between the input entered and the generated image. At this stage, processing occurs at the level of numerical information, and there is no graphic image processing. The latter occurs in the third block, where numerical representations are translated from a neural network with a decoding function into visually perceptible images [Rombach et al. 2022].

Related studies on AI applied to architectural design

Some studies related to AI in the field of architectural design are focused on highlighting the potential and limitations of the technology. In most cases, the potential is recognized as a support in the creative process [Jaruga-Rozdolska 2022; Paananen et al. 2023]. Among other potential uses, the ability to imagine abstract forms, reimagine biomimetic architecture, revisit traditional architecture, and visualize photorealistic advancements starting from architectural sketches are mentioned. The identified limitations are related to the possibility of control and customization of processes, insufficient consideration of structural feasibility aspects, and potential stylistic-architectural inconsistency in the generated results [Hegazy et al. 2023]. Case studies related to architectural projects where AI has been applied include the ideation phase, the generation of sketches with specific graphic styles, the addition of people and objects to existing images, the combination of various parts of images into a coherent composition, the variation of an initial image, the change of the graphic style of an existing image, floor plan design, exterior and interior design, texture creation, and urban planning [Ploenning et al. 2022; Yildirim 2022].
One educational study involves integrating AI techniques with traditional techniques in a first-year university design representation course, where the authors observed an improvement in students' interpretative and compositional abilities [Tong et al. 2023]. Students were asked to create a composition of solids and draw hand-drawn orthogonal projections and isometric axonometry. Then, they were required to generate a series of images using *Midjourney* with specific keywords. Finally, they were instructed to combine the two previous productions using various techniques.

Potential of text-to-image AI for architectural drawing

To experiment with the potential contribution of AI in the preliminary phase of the project, the moment when representation contributes to ideation and prefiguration, it was decided to work on idea definition and visualization.
Three different graphic inputs were hypothesized: two external perspective views of a three-dimensional volumetric model and a sketch of an interior, intentionally lacking characterizations except for the minimum necessary for spatial definition and framing. These graphic inputs, thanks to the *ControlNet* extension, are entrusted with incorporating the general morphological setting of the project into the generative process, while textual inputs are used to describe the desired graphic techniques and any architectural features related to materials, context, and additional stylistic characteristics that one wishes to include. The results of these initial experiments demonstrate the remarkable flexibility of AI in (re)creating different graphic techniques, ranging from pencil drawings to colored pencils to watercolors, with a significant ability to integrate both natural and artificial context elements. Simultaneously, the AI's addition of detailed elements such as textures, perforations, and materials contributes to the advancement of ideation in a process where it can be hypothesized that some of these elements may actually be inserted into the

Fig. 7. Images generated with StableDiffusion *to simulate various graphic techniques. From top to bottom: pencil, colored pencils, monochromatic watercolor, and colored watercolor. Prompt: linear, exterior view, contemporary architecture, highly detailed architecture, large windows, concrete, architectural drawings, technical drawings, [desired graphic technique], line drawings, working drawings, architectural sketches, conceptual style, abstract (authors' processing).*

Fig. 8. Images generated with StableDiffusion *for photorealistic visualizations. On the left, exterior views, prompt: exterior home view, concrete walls and roof, large glass windows, small rectangular swimming pool, garden, garden furniture, clouds. In the center, exterior views with curved surfaces, prompt: a pavilion in a contemporary architecture style, covered with reflective panels, in a square with people and trees. On the right, interior views, prompt: home interior view, modern architecture, large glass windows with curtains, timber framing, wood flooring, concrete ceiling, steel staircase, large sofa with pillows, armchair, coffee table with flowerpot, carpet, plants, lamp, minimalist style furniture, sunlight from windows, daylight (authors' processing).*

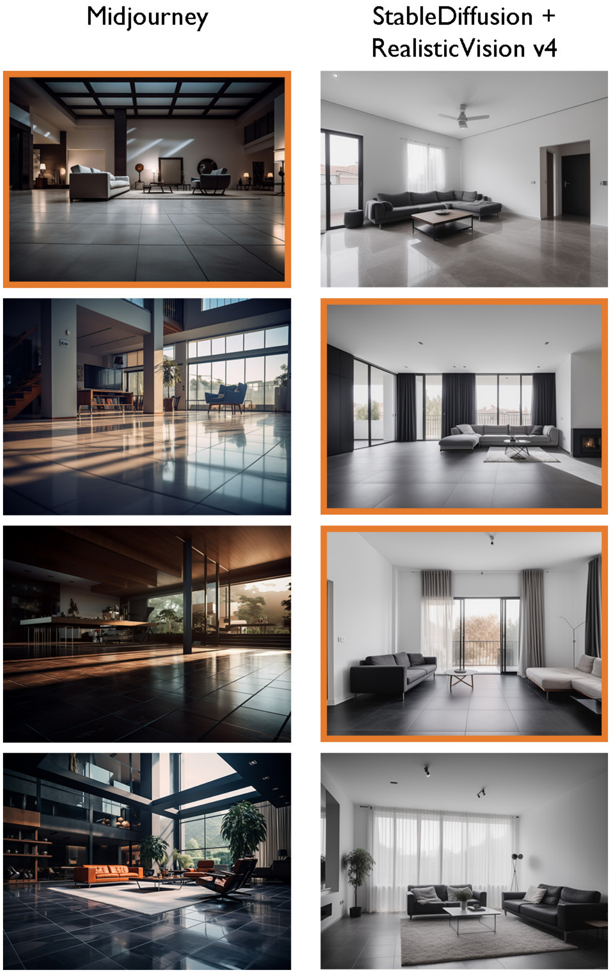


Fig. 9. Images generated through the prompt: central perspective, home interior view, floor with regular dark square tiles, modern architecture, minimalist style furniture, daylight (authors' processing).

ongoing project in a repeated human-machine exchange (fig. 7).
The potential contribution in terms of idea definition through the fast generation of variations is more evident when requesting the AI to produce photorealistic images. In this case, the ability to propose variations based on the textual prompt is more recognizable. Experiments conducted on external views demonstrate the variety of materials and interpretations of the simple volumetric schemes provided as input, as well as the skill in creating contextual settings (fig. 8). Similarly, experiments based on a digital sketch of an interior environment highlight the ability to combine colors and materials but also the inclination to add elements such as curtains and floor elevations. Minor elements such as light points and furnishings also appear. The distribution of these integrations generally aligns with the overall setting.

Limitations of text-to-image AI in representation

The limitations investigated in this paragraph [3] particularly concern the aspect of architectural representation (perspective, reflections, lighting/shadows).
At present, AI has no awareness of the projective rules underlying correct perspective construction. While this assertion could be deduced based on the theoretical principles behind the technology, it is also experimentally confirmed. By adding a descriptive part about the representation method (central perspective) [4] in the textual prompt, we arrive at results where central perspective is present only in some of the generated images (fig. 9). When analyzing the perspective setting of two of the previous generated images, it is observed that the vanishing lines of the square floor tiles (horizontal lines perpendicular to the picture plane) do not converge to a unique point (fig. 10). Additionally, by tracing diagonals from the two visible ends of the square tiles in the images, it can be noted that the intermediate intersections are not perfectly aligned with the diagonals. Therefore, the perspectives are perceptually effective but not correct as projections. The results of the perspective experimentation suggest that the AI has not been trained to correctly recognize different representation methods.

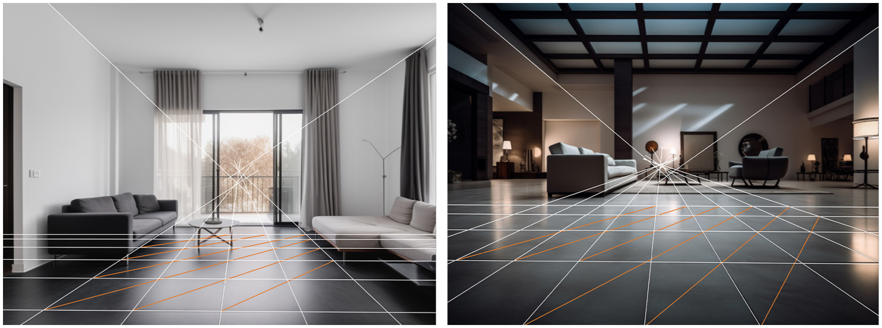


Fig. 10. Perspective analysis of two of the previous images generated with StableDiffusion *on the left and* Midjourney *on the right (authors' processing).*

Fig. 11. Images generated with StableDiffusion *on the left and* Midjourney *on the right through the prompt: home interior view, bedroom, modern architecture, minimalist style furniture, mirror with reflections (authors' processing).*

Fig. 12. Images generated with StableDiffusion *on the left and* Midjourney *on the right through the prompt: home interior view, modern architecture, minimalist style furniture, dramatic lighting and shadows (authors' processing).*

The analysis of elements reflected in mirrors revealed inconsistencies in some solutions, highlighting the spatial unawareness of AI. In particular, the mirror reflection lacks some elements present in the scene (such as the blanket on the bed in figure 11 on the left) or shows these elements in an inconsistent position (the same blanket in figure 11 on the right, which in the scene is draped over the sofa armrest, while in the reflection, it is spread out on the other side).
The study of shadows further confirms that the construction of images generated by AI lacks awareness of the three-dimensional space it represents. Very often, the light rays entering from windows produce shadows that are not consistent with the window fixtures (fig. 12).

Conclusions

The analysis of how text-to-image AIs function, the review of studies on the subject, and the experiments conducted allow us to outline initial reflections on AIs' possible role in architectural project representation.
The experiments highlight both potentials and weaknesses. Among the strengths, one can certainly include the high generation speed, which allows for high-quality visual images to be produced in just a few seconds, the flexibility of (re)producing graphic techniques, and the coherence with the provided prompts. These potentials make it possible for rapid shifts back and forth in the design process, from the initial ideation phase to the advanced visualization of the idea. Among the weaknesses, in addition to the previously discussed limitations in terms of representation accuracy, there are concerns raised by multiple parties regarding the legitimacy of the copyright of the methods used to create the training databases [5] and the presence of potential cultural biases induced in the AIs. For instance, it is worth noting how the squares proposed in figure 8 reflect North American models, diverging from the European conception of public space. Furthermore, this AI type is unsuitable for creating technical project documents. There is a substantial conceptual difference between artificial intelligence and human intelligence, stemming from the training and generation processes of AIs. The latter is interpolative intelligence, highly efficient at interpolating existing values within the training database and generating a value that is not present but is never entirely alien to it. They cannot extrapolate new values, not only those not present but entirely foreign to the database. This second form of intelligence is typically human [Del Campo 2022a]. The potential of human-AI co-creation in the ideation phase seems to lie precisely in the collaboration between two different types of intelligences, in which the interpolative one, initiated and guided by human inputs, proposes "Familiar but Strange" images [Del Campo 2022b, p. 28] from which human intelligence can draw inspiration for developing innovative ideas. This hypothesis renews the issue of authorship that algorithmic design/representation has already raised, leading to the idea that shared authorship among multiple agents (human or artificial) is inherent in the progress of the digital revolution in architecture. In this context, human authorship should still be understood as 'primary' because it plays the role of creating general rules, the Deleuzian 'objectiles', from which 'secondary' artificial authorships will derive individual forms, the 'objects' [Carpo 2011, p. 40, 123-128]. The transition from the architect as the designer of individual forms to that of a designer of general rules is already underway and has expanded the spectrum of adopted languages. Starting from the *second digital turn*, architects have learned to compose scripts and algorithms alongside graphic representation, and now, with the advent of AI based on prompts, they are called upon to integrate written natural language into their design methods.
This latest challenge posed by the digital revolution should stimulate reflection on representation languages in a broader sense and their teaching, which is one of the possible fields of interdisciplinary research in the present and near future of architectural drawing.

Credits

The authors have equally shared all phases of the research. For the purpose of drafting the article: M.F.M. wrote the *Introduction*, *Back and Foreward in Latent Space*, *Potential*, and *Conclusions*; S.M. wrote *(Recent) Past and Present*, *The Main Platforms*, *Related Studies*, and *Limitations*.

Notes

[1] Data as of July 2023, source <https://discord.com/servers> (accessed 24 July 2023).

[2] <https://rom1504.github.io/clip-retrieval/?back=https%3A%2F%2Fknn.laion.ai&index=laion5B-H-14&useMclip=false> (accessed 24 July 2023).

[3] The experimentation was conducted on two AI platforms: *Midjourney* and *StableDiffusion*, associated with an additional training model called *RealisticVision* v. 4.

[4] The prompt also included a specification regarding the presence of a floor composed of square tiles to enable subsequent perspective analysis.

[5] <https://www.egair.eu/> (accessed 24 July 2023).

Authors

Matteo Flavio Mancini, Department of Architecture, Roma Tre University, matteoflavio.mancini@uniroma3.it
Sofia Menconero, Department of History, Representation and Restoration of Architecture, Sapienza University of Rome, sofia.menconero@uniroma1.it

Reference List

Carpo, M. (2011). *The alphabet and the algorithm.* Cambridge - London: The MIT Press.

Carpo, M. (ed.). (2013). *The digital turn in architecture 1992-2012*. Chichester: John Wiley & Sons.

Carpo, M. (2017). *The second digital turn: design beyond intelligence*. Cambridge - London: The MIT Press.

Colton, S. et al. (2021). Generative Search Engines: Initial Experiments. In A. Gómez de Silva Garza et al. (Eds.). *Proceedings of the 12th International Conference on Computational Creativity*, Mexico City, 14-18 September 2021, pp. 237-246. Mexico City: ACC.

Crowson, K. et al. (2022). *VQGAN-CLIP*: Open Domain Image Generation and Editing with Natural Language Guidance. In *arXiv*. <https://arxiv.org/abs/2204.08583> (accessed 18 July 2023).

Del Campo, M. (2022a). When Robots Dreams. In Conversation with Alexandra Carlson. In *Architectural Design*, No. 03, V. 92, pp. 47-53.

Del Campo, M. (2022b). *Neural Architecture. Design and Artificial Intelligence*. Novato: Oro Editions.

Dhariwal, P., Nichol, A. (2021). Diffusion Models beat GANs on Image Synthesis. In M. Ranzato et al. (eds.). *Advances in Neural Information Processing Systems*, V. 34, pp. 1-15. Cambridge: MIT Press.

Goodfellow, I. et al. (2014). Generative Adversarial Nets. In Z. Ghahramani et al. (eds.). *Advances in Neural Information Processing Systems*, v. 29, pp. 1-9. Cambridge: MIT Press.

Hegazy, M., Saleh, A.M. (2023). Evolution of AI role in architectural design: between parametric exploration and machine hallucination. In *MSA Engineering Journal*, V. 2, No. 2, pp. 262-288. <www.doi.org/10.21608/MSAENG.2023.291873> (accessed 18 July 2023).

Hong, W. et al. (2022). *CogVideo*: Large-scale Pretraining for Text-to-Video Generation via Transformers. In *arXiv*. <https://arxiv.org/abs/2205.15868> (accessed 18 July 2023).

Jaruga-Rozdolska, A. (2022). Artificial intelligence as part of future practices in the architect's work: MidJourney generative tool as part of a process of creating an architectural form. In *Architectus*, V. 3, No. 71, pp. 95-104.

Jun, H., Nichol, A. (2023). *Shape-E*: Generating Conditional 3D Implicit Functions. In *arXiv*. <https://arxiv.org/abs/2305.02463> (accessed 18 July 2023).

Nichol. A. et al. (2022). *Point-E*: A System for Generating 3D Point Clouds from Complex Prompts. In *arXiv*. <https://arxiv.org/abs/2212.08751> (accessed 18 July 2023).

Paananen, V. et al. (2023). Using Text-to-Image Generation for Architectural Design Ideation. In *arXiv*. <https://arxiv.org/abs/2304.10182> (accessed 18 July 2023).

Ploennings, J., Berger, M. (2022). AI Art in Architecture. In *arXiv*. <https://arxiv.org/abs/2212.09399> (accessed 18 July 2023).

Prix, W. et al. (2022). The Legacy Sketch Machine. From Artificial to Architectural Intelligence. In *AD, Machine Hallucinations: Architecture and Artificial Intelligence*, V. 92, No. 3, pp. 14-21.

Ramesh, A. et al. (5 January 2021). DALL-E: Creating images from text. <https://openai.com/research/dall-e> (accessed 18 July 2023).

Radford, A. et al. (2021). Learning Transferable Visual Models from Natural Language Supervision. In M. Meila, T. Zhang (eds.). *Proceedings of the 38th International Conference on Machine Learning*. Virtuale, 18-24 July, V. 139, pp. 8748-8763. Maastricht: ML Research Press.

Reed, S. et al. (2016). Generative Adversarial Text to Image Synthesis. In M. F. Balcan, K. O. Weinberger (a cura di). *Proceedings of the 33rd International Conference on Machine Learning*, V. 48, pp. 1060-1069. Maastricht: ML Research Press.

Rombach, R. et al. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, 18-24 June, pp. 10674-10685. New York: IEEE.

Singer, U. et al. (2022). *Make-A-Video*: Text-to-Video Generation without Text-Video data. In *arXiv*. <https://arxiv.org/abs/2209.14792> (accessed 18 July 2023).

Tong, H. et al. (2023), An attempt to integrate AI-based techniques into first year design representation course. In K. Vaes, J. Verlinden (a cura di). *Connectivity and Creativity in times of Conflicts. Cumulus Conference Proceedings*. Anversa, 12-15 April, pp. 1-5. Antwerp: University of Antwerp.

Tsigkari, M. et al. (29 March 2021). Towards Artificial Intelligence in Architecture: How machine learning can change the way we approach design. In *Plus Journal*, <https://www.fosterandpartners.com/insights/plus-journal/towards-artificial-intelligence-in-architecture-how-machine-learning-can-change-the-way-we-approach-design> (accessed 18 July 2023).

Wallish, S. (2022). GAN Hadid. In S. Carta (ed.). *Machine Learning and the City: Applications in Architecture and Urban Design*, pp. 477-481. Hoboken-Chichester: John Wiley & Sons.

Yildirim, E. (2022). Text-to-image generation A.I. in architecture. In H. Hale Kozlu (ed.). *Art and Architecture: Theory, Practice and Experience*, pp. 97-119. Lyon: Livre de Lyon.

Zhang, L., Agrawala, M. (2023). Adding Conditional Control to Text-to-Image Diffusion Models. <https://arxiv.org/abs/2302.05543> (accessed 18 July 2023).