Journal of Building Engineering 80 (2023) 107999

Contents lists available at ScienceDirect

Journal of Building Engineering

journal homepage: www.elsevier.com/locate/jobe

Artificial intelligence's effects on design process creativity: "A study on used A.I. Text-to-Image in architecture"

Nervana Osama Hanafy

Department of Architecture Engineering, Modern University for Technology & Information, Mokattam, City, 11571, Egypt

A R T I C L E I N F O

Keywords:
Artificial intelligence
Image in architecture
Text-to-Image
Natural language processing
Computer vision

A B S T R A C T

The field of artificial intelligence has introduced a groundbreaking process called text-to-image generation, which allows humans to visualize ideas in new ways. This study suggests that architects and designers can utilize these tools in conceptual design to enhance their creativity. AI provides designers with fresh perspectives and opens up new possibilities for design concepts. These tools also allow for easier manipulation of data and the development of innovative solutions to complex problems, expanding the designer's range of creative abilities beyond traditional methods. We are examining the application of usage-based models to different activities, exploring the suitability of DALL-E 2, Midjourney, and StableDi platforms for typical examples in architectural design. By using NLP techniques to extract typical usage patterns from a data collection of 40 million Midjourney inquiries, we can analyze these patterns and how they are currently being used. We examined architectural glossaries to identify keywords related to architecture. We only included keywords that appeared at least 10 % of the time, and the names of 941 famous architects from Wikipedia were added to the keyword list. This allowed us to identify 2.6 million searches (6.8 %) that potentially had an architectural intention, including 1 million (2.6 %) that specifically mentioned "architect", "interior", or "exterior" design. Our filters ensured the accuracy of our findings. Conclusions, the use of technology can greatly enhance the engineering design process. Architects and other creatives find it particularly exciting and useful due to its user-friendly interface and ability to generate a wide range of images.

Abbreviations

AI	Artificial Intelligence
ML	Machine Learning
IoT	Internet Of Things
AIAD	Artificial Intelligence Assisted Design
NLP	Natural Language Processing
GAN	Generative Adversarial Network

E-mail address: Nervana.osama@outlook.com.

https://doi.org/10.1016/j.jobe.2023.107999
Received 2 June 2023; Received in revised form 26 September 2023; Accepted 21 October 2023
Available online 30 October 2023
2352-7102/

1. Introduction

The way we work and interact with one another has changed and improved significantly as a result of the development of digital tools over the past century and the appearance of highly technological advancements like artificial intelligence (AI), machine learning (ML), the internet of things (IoT), and digital twins. Massive volumes of data may now be consumed and analyzed more quickly than ever before because to these advancements. Assume that computer and artificial intelligence (AI) development will proceed at the same rate as it has for the previous 50 years (see Fig. 1 and 3 and 4).

Even if it will be some time before we encounter an AI with traits similar to those of humans, the AI we presently have needs to be addressed in how it will be developed in the future because it may be projected to have a big impact on our society. While waiting, we can investigate how the use of AI affects human traits like creativity, which this paper will do from the standpoint of architecture. Given that creativity is a very complex human quality with numerous interrelated factors at play and one that is crucial to design thinking, the impact AI has on creativity becomes particularly intriguing. Digital platforms and tools have been considered as aiding design thinking since their introduction in the early 1960s through simulation and modelling of design options. They are also considered to be instruments for improving cognitive processes and skills, such as creativity [1].

According to the McKinsey Digital Globalization Index, the construction industry is one of the least digitized industries globally. In response, fields related to construction, such as architecture, engineering, and construction, are embracing digitization and intelligence to improve automation, productivity, and reliability. To achieve digital strategies in construction-related fields, AI serves as an efficient and feasible solution to change traditional construction project execution modes. Data in construction-related fields are stored in various electronic text formats, such as Word, Sheet, Email, XML, HTML, PDF, CAD, IFC, etc. These formats contain human language, and NLP provides an intelligent way to process text data, allowing the intelligent agent to learn from human language and automatically complete knowledge representation, retrieval, and reasoning processes in a human-like way. Drawing tools have

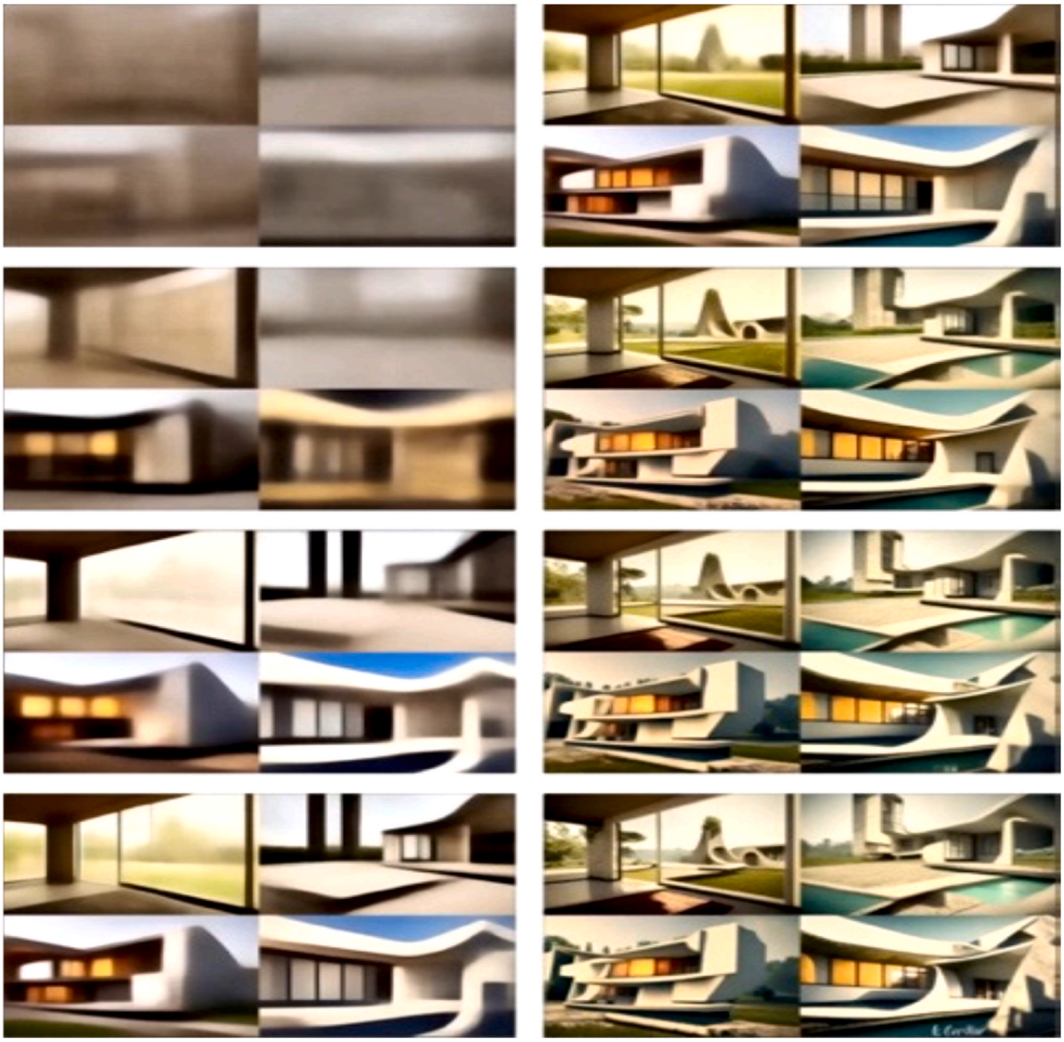


Fig. 1. Sequences of the diffusion modeling-created image [22].

Fig. 2. Animation of text diff used to transform a Victorian house (Fig. 2) into a modern one. The transformation is determined by the captions "a Victorian house", which describes the architecture of the house, and "a modern house", which describes how the architecture of the house should be changed. house.gif [25].

advanced significantly since the introduction of digital tools like CAD and CAAD in the early 1960s, thus it is crucial to look into the potential effects that cutting-edge technologies like AI may have on the creative design process. In addition to being a relatively new technology, creativity, one of the more complex human abilities, is also challenged by it. Therefore, NLP technologies are essential for achieving further intelligence in construction-related fields.

There are more and more NLP-related studies being carried out in construction-related areas. NLP is applied in document management [2,3], safety management [4,5], compliance checking [6], risk management [7–9], and Building Information Modeling (BIM) [10]. However, there is no relevant review research referring to NLP applications in construction-related areas. This lack of literature review makes it difficult for researchers who want to enter a new area. Although most review studies in construction-related fields focus on scient metric analysis or summary of main ideas, they often overlook the importance of sorting and summarizing datasets/data sources, technologies, and tools. For example, Martinez et al. [11] conducted a comprehensive scient metric analysis on computer vision applications in the construction field, discussing current research status and future trends without summarizing data and technology. Fang et al. [12] summarized the application of computer vision-related technology for behavior-based safety in construction, ignoring scient metric analysis and the review of data used in relevant literature. Although Yan et al. [13] reviewed data sources, algorithms, and technologies related to data mining, the summary of data sources is only a simple statistic of data acquisition methods without providing links to data sources and other related information. Romero-Silva and de Leeuw [14] conducted a review work of text mining analysis in Operations Research and Management Science (OR/MS) subject area, which focused only on publication analysis and future development. An all-around review is necessary to provide readers with an overall in-depth understanding of the research field.

On the other hand, it is now possible to produce visuals from provided text thanks to the advancement of AI algorithms. The technique of generating an image from text is known as text-to-image generation. Since the images that are produced are frequently very realistic, this method can be used to provide a wide variety of visual content. The text-to-image generating technique is projected to become a more crucial tool in the field of architecture as AI technology develops [15]. As a shift of algorithms from 2D to 3D generation, it might also be utilized to create interactive, immersive experiences for people to examine a variety of design options [16]. Through text-to-image generation, artificial intelligence is revolutionizing the way humans visualize ideas. Tools like DALL-E 2 and Midjourney have made it possible for AI to create a wide range of visually appealing works. For use in conceptual design, these tools will be put to the test in the real world by architects and designers. AI enables designers to investigate a greater variety of design concepts from fresh angles. By increasing the chance of serendipity, these systems broaden the range of creative abilities beyond what can be achieved through more conventional means. They make it easier to manipulate data and come up with novel solutions to challenging problems.

Modern art systems can convert text descriptions into appealing visuals, providing powerful tools for conceptual design. This study examines the practicality of these systems in architectural design scenarios, utilizing Natural Language Processing (NLP) to analyze usage patterns and integrating these insights into an architectural design project. The notion of the "Interaction between Creativity and AI Generation Model" stands as a highly contentious subject. By its very essence, it could also be framed as a qualitative research subject. The underlying query (Can AI exhibit creativity akin to humans?) occasionally engenders unnecessary debates, making it imperative to elucidate the scope of creativity within the confines of this paper, and to quantify/classify the subject matter. This study first examines the connection between architecture and artificial intelligence. An introduction to the artificial intelligence platforms Dall-E 2, Midjourney, DiffusionBee, and Motion Leap is given after a brief discussion of the techniques needed to generate text to image. The architectural, interior, and urban design come last.

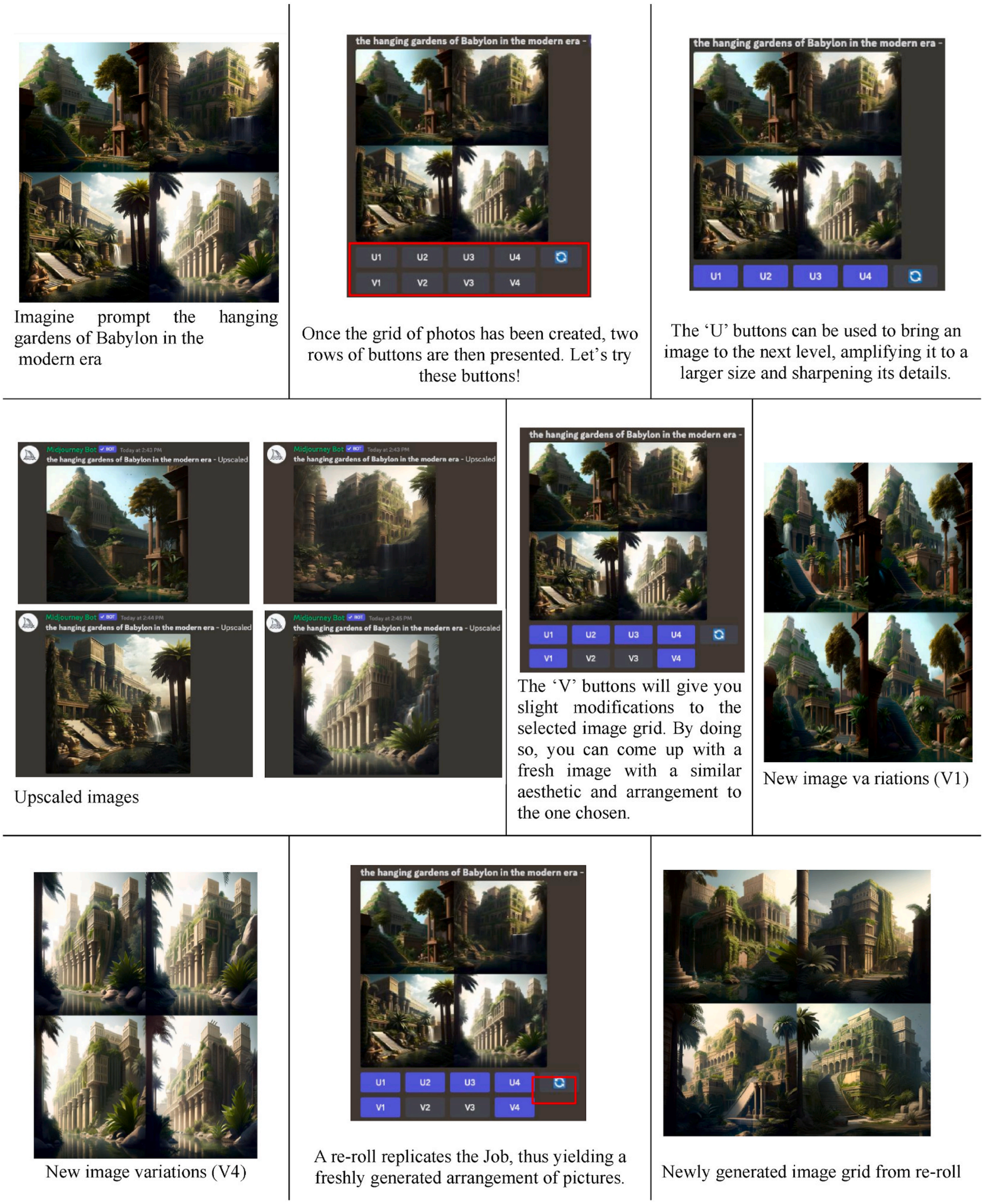


Fig. 3. Text-to-image prompt example Use of Midjourney [28].

1.1. *The aim of this study*

- This paper addresses a comparative examination of an overarching generative AI platform and concentrates on evaluating the influence of the AI's 'creative' abilities on the process.
- This study suggests that architects and designers can utilize these tools in conceptual design to enhance their creativity.
- AI provides designers with fresh perspectives and opens up new possibilities for design concepts.
- These tools (DALL-E 2, Midjourney, and StableDi) also allow for easier manipulation of data and the development of

Innovative solutions to complex problems, expanding the designer's range of creative abilities beyond traditional methods.

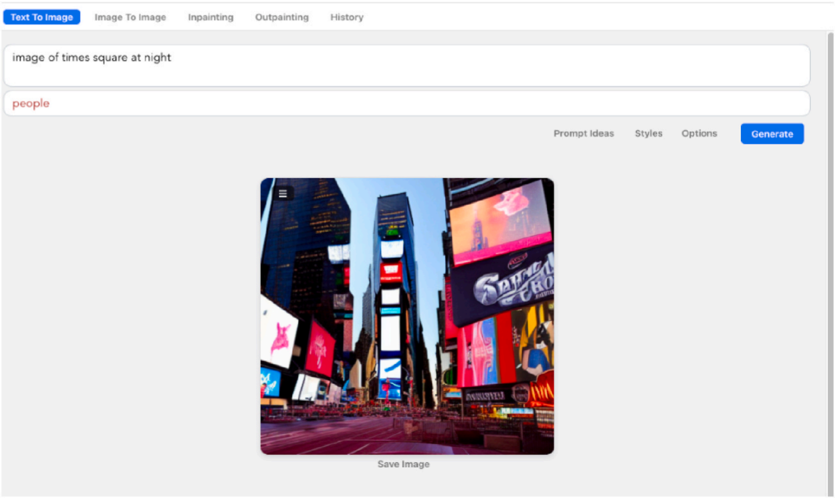


Fig. 4. User interface of DiffusionBee [29].

2. Material and methods

2.1. Text to image generation methods

The term "Text-to-Image Generation" describes computer-based methods that may turn human-written textual descriptions, such as keywords or phrases, into visually represented concepts having the same semantic meaning as the text. To find the best feasible match between text and images, early image synthesis researchers used supervised methods and text-to-image correlation analysis. Images used in applications like image classification frequently have just one thing in them that needs to be labelled. This challenge may get even more challenging as computers become increasingly capable of comprehending complex scenarios. Making captions for pictures is one of these responsibilities. This position is tough for two reasons. In order to produce a caption that is semantically understandable and syntactically fluid, the system must first represent language and common-sense knowledge in addition to object recognition. As a result, the system is better able to identify important semantic pieces within an image, understand how those elements are related, and produce a coherent description of the image's overall content. Additionally, the complexity of the scenes in the picture makes it impossible for the simple category attribute to adequately capture all of the minor differences between them [17]. The real turning point came in 2014 with the development of generative adversarial networks (GANs) by computer scientist Ian Goodfellow [18].

The process is driven by two antagonist neural networks. New images are created by the first, known as the generator, and are contrasted with a training set of samples by the second, known as the discriminator. Comparable to showing a fake piece of art to a critic in place of the original. The discriminator will keep rejecting output images if their quality is worse than that of the input images. After training is over, the discriminator can be turned off while the generator keeps producing output of a high caliber. The work of Refik Anadol, which employs various GAN types, is an intriguing example of this technique in action [19].

One of the more recent achievements in the field of machine learning is image recognition, sometimes known as machine vision. Visual "hallucination" is a relatively new phenomenon, but [20]. Alexander Mordvintsev, a Google developer, made the ground-breaking finding that a neural network could be trained to work in reverse in 2015. It can now produce fake graphics and identify real ones. This allowed him to produce the psychedelic computer vision program Deep Dream, whose effects sparked a backlash in the artistic community.

The most modern and effective method for creating images is diffusion modelling. A thorough understanding of mathematics is necessary for the diffusion model. Simply put, you start with an image and keep adding noise to it without ever looking back.

Every new step is regarded as the start. These procedures can all be carried out separately from one another and without using any previous photographs as a guide. The resulting image is completely random noise and lacks any originality. The diffusion model is said to be the best picture synthesis method by Dhariwal and Nichol [21]. AI can already "dream" in the visual realm because to improvements in hardware for processing enormous datasets and machine learning algorithms.

2.1.1. Dall-E 2

DALL-E 2, a sizable AI language model, was produced by OpenAI. It can produce graphics from written descriptions by combining natural language processing and computer vision algorithms [23]. One of DALL-E 2's distinguishing features is its capacity to produce a wide range of images, from photorealistic representations of objects and settings to more abstract, stylized images. Because of this, it works well for creative projects like design, illustration, and art.

Users can easily generate graphics using DALL-E 2's simple and user-friendly interface based on their own text descriptions. A textual description of the desired image is all that is required to use DALL-E 2, and the AI model will create the related image. The model can then receive feedback from the user and utilize it to evaluate the created image, gradually learning and developing.

A remarkable text-to-image AI called Dall-E 2 can produce incredibly lifelike graphics from word descriptions. Dall-E 2's user interface is made to be both simple and effective, allowing users to rapidly produce beautiful visuals from their written descriptions. Dall-E 2's user interface is notable for its simplicity. The interface walks users through the process of producing their own photographs using simple structure and directions. This ease of use makes it possible for anyone with little to no expertise using A.I. technology to swiftly and easily produce gorgeous photos [24].

The versatility of the Dall-E 2 user interface is a key feature. Users can alter their photographs using the interface in a number of ways, such as changing the colour scheme and viewpoint. Users can produce photos that are one-of-a-kind and customised to their individual requirements and vision thanks to this versatility.

However, Dall-E 2's user interface's versatility in terms of accepting different types of inputs is arguably its most amazing feature. Dall-E 2's AI technology can interpret text descriptions of various lengths and complexity, enabling users to produce images from even the most intricate and difficult text descriptions. In conclusion, Dall-E 2's user interface is a crucial part of the technology's strength and worth.

2.1.2. Midjourney

Users can enter content and have a text-to-picture AI program called MidJourney produce an image in response. This technology offers a wide range of possible uses, from helping designers and artists quickly draw out concepts to helping people with visual impairments understand printed content [26].

For MidJourney, the arts and design are a potential application. MidJourney enables users to produce visuals based on their written descriptions, which can help designers and artists quickly sketch out ideas and thoughts. This can be especially helpful for those who want to explore their creative ideas but lack a strong grounding in traditional art approaches.

The interface's complexity and the prompting process are two of MidJourney's standout features. To create an image, a user merely needs to enter the required text into the programme and click the "generate" button. The input text is then analyzed by MidJourney using sophisticated machine learning methods to produce a corresponding image. Access to Midjourney is made possible by the online chat programme Discord. A user will only be able to create a certain amount of photographs using the Midjourney tool before having to subscribe. The/imagine command and additional queries (variations, upscales) can be executed by the user around 25 times for free. The terms "jobs" and "GPU-minutes" are also used to describe them. U1, U2, U3, and U4 are the four buttons in the top row that can be used to enlarge the selected image. When an image is upscaled, a larger version of the original with roughly 1024x1024 pixels is produced. By default, this will produce additional information. The V1, V2, V3, and V4 buttons on the bottom row let you make variations of the chosen image [27].

Four additional photographs that are conceptually and formally comparable to the chosen image are produced when variants are created. To save the image to your local computer, expand it with your mouse by clicking, then choose Save image from the context menu.

2.1.3. Diffusion Bee

A cutting-edge text-to-image AI programme called DiffusionBee enables users to generate photorealistic images from word descriptions. High-resolution photographs created by the programme can be utilized for a range of tasks, such as social media posts, advertising, and creative expression.

DiffusionBee stands out for having an intuitive interface. Even for individuals who are not familiar with text-to-image artificial intelligence technologies, the programme is meant to be simple to use and straightforward. Users merely need to give a brief description of the desired scene in order to generate an image. The programme will quickly produce a photorealistic image using this description.

One of DiffusionBee's most crucial applications is advertising. The programme enables companies to easily generate visually beautiful commercials without the need for expensive photography or graphic design expertise. By doing this, businesses can create advertisements that are sure to grab the attention of potential customers while saving time and money.

The uses of DiffusionBee go beyond advertising and encompass blog postings, social media updates, and other types of digital material. The programme can assist users in enhancing their online profile and producing more aesthetically attractive material by giving them a quick and easy way to make photographs with a professional appearance.

DiffusionBee is a text-to-image AI programme that works well and is simple to use. It offers both organizations and people a number of advantages. The programme makes it possible to create outstanding visual content in a few easy clicks because to its powerful algorithms and user-friendly interface.

2.1.4. Motionleap

The text-to-image AI in MotionLeap, a mobile image-editing application, has gained widespread adoption. The user-friendly interface of MotionLeap makes it simple for users to create high-quality graphics from text descriptions, making it a vital tool for bringing concepts to life.

MotionLeap's user interface is distinguished by its simplicity. Users with little to no expertise using A.I. technology may make beautiful photographs fast and simply while on the road because to its straightforward design and ease of usage. To accomplish this simplicity, customers are guided through the process of making their own photographs using precise, succinct instructions and

a simple structure.

The user interface of Motion Leap must be adaptable. Users can alter their photographs using the interface in a number of different ways, such as by changing the theme or the image's style. Users can produce photos that are one-of-a-kind and customized to their individual requirements and vision thanks to this versatility.

The most striking aspect of Motion Leap, though, might be how quickly it loads. MotionLeap's AI technology makes it possible for users to quickly and simply bring their ideas to life by processing written descriptions and producing accompanying visuals in a matter of seconds.

In conclusion, a key element of the technology's strength and value is the user interface of MotionLeap. MotionLeap is a vital tool for anyone looking to bring their ideas to life with portability because of its user interface's simplicity, versatility, and speed, which allows users to easily produce spectacular visuals from written descriptions.

2.2. Tools for comparing text and images

Dall-E, Midjourney, Diffusion Bee, and Motion Leap are cutting-edge AI technologies with outstanding skills in the area of text-to-image generation. Despite the high level of development shared by all four AIs, they all differ greatly from one another.

The way the four AIs handle text descriptions is one of the most important differences between them. A deep learning neural network is used by Dall-E, a more traditional machine learning strategy is used by Midjourney, a cutting-edge diffusion-based algorithm is used by Diffusion Bee, and a combination of neural network and machine learning methods is used by MotionLeap. This suggests that each artificial intelligence (A.I.) when creating visuals from text descriptions has its own advantages and disadvantages.

Another significant difference between the four AIs is their user interfaces. Those with little to no expertise using AI technology may quickly and easily generate photographs with Dall-E thanks to its user interface, which is made to be simple and straightforward.

Midjourney's user interface, on the other hand, is more complicated and might be trickier to utilize. Users can fine-tune their photographs using Diffusion Bee's fully customizable user interface to suit their own requirements and preferences.

Additionally, the customizable and user-friendly Motion Leap user interface makes it simple for mobile device users to create high-quality photos from text descriptions.

With the exception of Diffusion Bee, all three AIs can quickly produce visuals from text descriptions. Dall-E, however, has the advantage of being able to handle a wider variety of input, including text descriptions of various lengths and complexity. As a result, Dall-E is a more adaptable tool for users who need to produce a variety of images from different text descriptions.

All in all, Dall-E, Midjourney, Diffusion Bee, and Motion Leap are exceptionally sophisticated artificial intelligence (AI) tools that can produce incredibly lifelike visuals from textual descriptions. Despite their major distinctions, all four artificial intelligences are useful tools for anyone trying to realize their ideas (see Fig. 5).

Both the internal and interface models' structures have an impact on how these models are used. We compare and contrast some of the features of DALL-E 2 with those of StableDi and Midjourney in Fig. 6. Similar to how they were in earlier technologies, the fundamental building components of basic technologies are greyed out. The distinctions are in the workflow they offer, which is frequently the outcome of their various interface strategies.

Fig. 6 makes clear that while the grey core work is comparable, the approaches used to provide the results are substantially dissimilar. Only a sequential scaling of resolution is performed by Midjourney, and it does so at multiple different sizes. Due to the restricted ability to remix the original text vector, mid-flight operations concentrate on developing and evaluating picture variants from several txt2img models, then down sampling the best ones.

In contrast, direct editing of loaded images or txt2img output is possible when using both DALL-E and StableDi. It does not include conventional image editing tools like stamping, layering, coloring, or sketching. Instead, any image manipulation must be performed via img2img-based procedures like stretching the canvas, clearing partitions, or a combination of the two. Every grid has the ability to

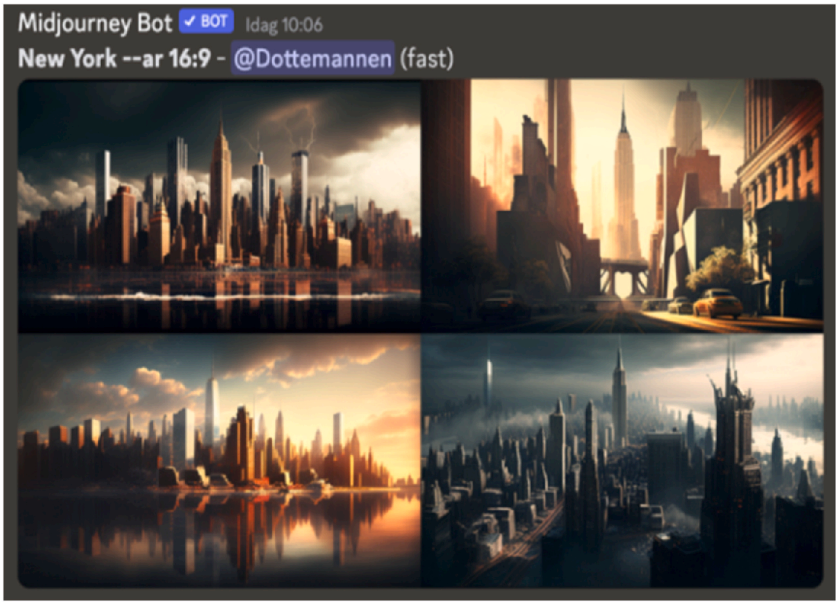


Fig. 5. Text-to-image AI module of MotionLeap's user interface [23].

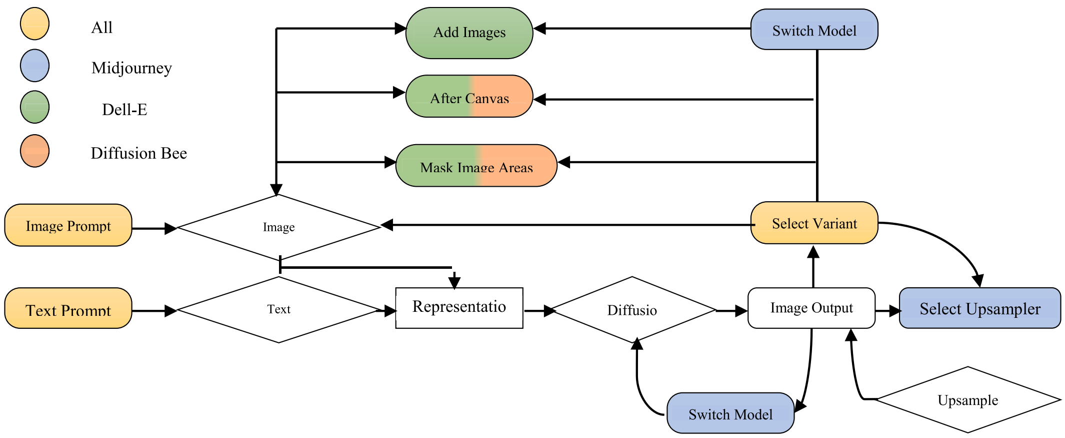


Fig. 6. Distinct models have distinct model geometry and image production techniques. Grey elements represent standard model building blocks, whereas colored elements represent user interaction points.

Table 1
Demonstrates the outcomes for the use scenarios.

Model	Txt2img	Img2img	In/Out-paint	Editing	Upscaling	Semantics
DALL-E	♦	♦	♦	◇	◇	◇
Midjourney	♦	◓	◇	◇	♦	◇
Stable Diffusion	♦	♦	◓	◇	◇	◇
Use Case						
Ideation	♦/♦	◒/♦	◇/◒	◇/◇	◇/◓	◇/◇
Sketches	◓/◓	◒/◓	◒/◒	◇/◇	◇/◓	◒/◇
Collages	◇/◇	◒/◇	◓/◓	♦/◇	◇/◇	◇/◇
Image Combination	◇/◇	◓/◇	♦/♦	◒/◇	◒/◇	◇/◇
Build Variants	◇/◇	◓/◒	♦/◓	♦/◇	◇/◇	◓/◇
Style Variants	◇/◇	♦/◓	◇/◇	◇/◇	◒/◒	◇/◇
Construction Plans	◓/◇	◇/◇	◇/◇	◓/◇	◓/◇	♦/◇
Exterior Design	◓/◓	◒/◒	◓/◓	◓/◇	◒/◒	◒/◇
Interior Design	◓/◓	◒/◒	◓/◓	◓/◇	◒/◒	◒/◇
Creating Textures	◓/◓	◇/◒	◇/◇	◒/◇	♦/◓	◇/◇

Table 1: Top part – Comparison of platforms with their supported features with: ♦full, ◓ limited, ◒bad, or ◇ no support, Lower part – Mapping of architectural use cases to features with: ♦high, ◓ some, ◒ low, or ◇ no importance; versus (/) how well it works: ♦ well, ◓somewhat, ◒ a little, ◇ not at all.

generate images with various sizes and aspect ratios. This may be done by either defining the size in the query or by editing the image later using an external paint programme.

Most importantly, all templates have quick image creation times that are quick enough to be used during creative sessions with clients or alone. Additionally, all three models in some way support the input of outside photos. As a result, they are simple to incorporate into composite operations.

The data that were used to train each model is one aspect that is absent from Fig. 6. The training data utilized in DALL-E and Midjourney is not well documented. However, the LAION-5B dataset [30], which is based on image and text data downloaded from the internet, was used to train StableDi. It is highly likely that Midjourney and DALL-E came from similar Internet datasets. However, it is evident that these models, which produced widely distinct visual patterns, were either biased by the training data or the training procedure. It is simple to produce both realistic and drawn output with DALL-E and StableDi.

Midjourney tends to have an impressionistic aesthetic. Target keywords can modify this, but even results that are blatantly realistic in Midjourney tend to be less concrete until a recently introduced form variable is expressly used on an earlier version of an image. Most patterns can be targeted with carefully chosen inputs, but for convenience, it is frequently preferable to use a platform that is proficient in producing the desired pattern.

2.3. Architectural use cases

Given the variations in technology and user interface that have been described, the platforms' supported architectural use cases vary greatly. We gathered a number of use scenarios where architects typically make or change photos in order to analyze this. The style differences across the platforms and the use cases make it difficult to directly compare them. We found that, in the majority of cases, the main distinguishing factor is whether or not a platform provides a particular technical feature that is necessary to realize the use case.

Therefore, we assessed the functionality needed for each use case and how well this already functions on supported platforms in terms of quality. We additionally map structured knowledge that goes beyond conventional image training datasets to show which platforms completely or partially support which characteristic. The outcomes for the use cases that we go over following are shown in Table 1.

- **Ideation:** The process of creating ideas by producing images at random. Txt2img models were created for this purpose, and it works great. Style and object references can be added with additional image prompts.
- **Sketches:** Creating architectural diagrams with a certain style and object in mind. This works effectively for typical training data instances, but less so for unique queries.
- **Collages:** Combining and adding people and things to existing photographs to give them life. It is possible to accomplish this by inpainting for specific items, but not for general requests like "Add many people."
- **Image Combination**: Putting together several image components that already exist (for instance, various buildings) on a canvas to produce a cohesive composite image.
- **Build Variants:** Producing variants of an existing drawing or image in which specific parts have been changed (such as the addition of a garage). This effectively utilizes inpainting.
- **Style Variants:** Taking an existing image and modifying its style without affecting the content (for example, from a sketch to photorealistic art deco). Certain models respond favourably to this.
- **Construction Plans:** Producing thorough floor plans that define spatial relationships. Since the model does not comprehend the semantics of line styles, areas, etc., this rarely works.
- **Exterior Design:** Finding a building's style and mood in relation to its surroundings and terrain. This is effective in typical situations.
- **Interior design**: Choosing a look or atmosphere for a room's interior. This is also effective in many situations.
- **Texturing:** Producing tiled designs to use as the surface materials for 2D or 3D models. Midjourney presently has no other features like this.

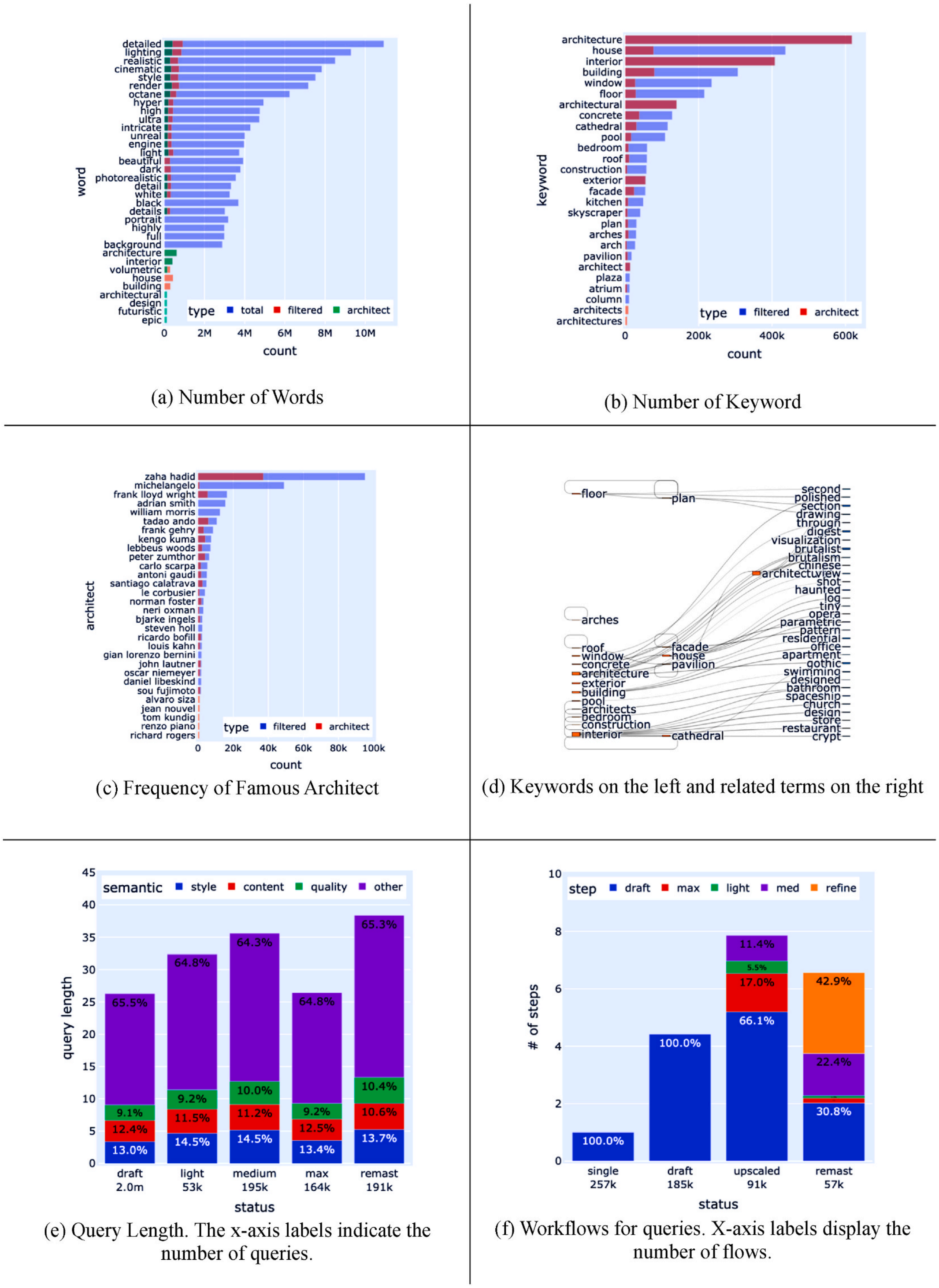
Although Stable Diffusion appears to support fewer features than DALL-E, its fundamental benefit—the ability to run it locally and provide more training data to the model to expose it to new architectural concepts—cannot be understated. This makes Stable Diffusion the most effective of the three models to be specialized on architectural designs, in addition to the high quality of its outputs.

2.4. Analysis of architectural queries

We also looked into the practical applications of these AI art systems. As Midjourney is the sole system for which a large number of user inquiries are publicly visible, we examined around 40 million queries made by Midjourney users. Since the main interface of Midjourney is the messaging software Discord, we were able to keep an eye on the public channels for questions that we deemed to be of an architectural character. We chose searches using the terms "architect", "interior", "exterior" design, or one of 38 keywords related to architecture, such as "building", "facade", or "construction". By only choosing terms from architectural glossaries that appeared in the queries alongside "architect", "interior" or "exterior" at least 10 % of the time, we were able to identify these keywords. As we saw that numerous inquiries truly refer to the style of these architects, we also added the names of 941 illustrious architects from Wikipedia to the list of keywords. By using these filters, we were able to pinpoint 2.6 million searches (6.8 %) that may have had an architectural intention, including 1 million (2.6 %) that had the words "architect", "interior", or "exterior" design in them specifically.

The following steps involve removing stop words from these searches and creating a Word2Vec model [31] to create a model of the occurrence and co-occurrence of keywords. Knowing that most Midjourney users do not construct whole phrases, but rather a combination of terms that pertain to the content, style, or render quality of the targeted image, is crucial to comprehending the results.

The most common terms used in the filtered queries are displayed in Fig. 7 (a). The frequency in blue throughout all 40 million queries, red within the 2.6 million filtered queries, and green within 1 million questions expressly involving "architect", "interior", or "exterior" design are shown by the three colors. All three classes have a similar frequency of the top 10. These frequently make



(a) Number of Words

(b) Number of Keyword

(c) Frequency of Famous Architect

(d) Keywords on the left and related terms on the right

(e) Query Length. The x-axis labels indicate the number of queries.

(f) Workflows for queries. X-axis labels display the number of flows.

Fig. 7. Visualizes the main results of our analysis.

reference to Midjourney-style commands like "detailed", "realistic", "cinematic", and "render". However, while some phrases, like "black," "full," or "portrait," are used frequently generally, they are rarely used in relation to architecture. As they are on our keyword list, other terms like "architecture," "interior," "house," and "building" only appear in our filtered results.

These keywords and their respective frequency are listed in Fig. 7(b). These keywords are filtered, thus since their overall frequency is the same as the filtered frequency, we do not display it. The keywords "architecture" and "interior" are the most and third most frequently used, respectively.

The frequency of well-known architects, as determined from 263.041 questions mentioning at least one of them, is shown in Fig. 7 (c). Given her instantly recognisable style and the potential popularity of her projects among the same social groups interested in experimenting with AI techniques, Zaha Hadid is by far the most commonly questioned architect. The second is Michelangelo, who, like William Morris, is frequently cited for his artistic rather than his architectural accomplishments. Top 10 names that are frequently used in expressly architectural contexts include Frank Lloyd Wright, Tadao Ando, Frank Gehry, Lebbeus Woods, Kengo Kuma, Peter Zumthor, and Antoni Gaudi.

The connections between keywords and the most likely associated phrase are displayed in Fig. 7(d). In order to analyze this, we used Word2Vec to forecast the most probable co-located term for each keyword on the left, weighted by probability. Here, interesting combinations include those including an inner apartment, a floor plan, parametric design, gothic arches, or a pavilion roof. From this, an auto-complete method for architectural inquiries can be created.

The average length of queries is shown in Fig. 7(e) depending on whether they were upscaled, remastered, or left in draught mode. Users typically upscale or remaster draught mode images if they prefer one of the varieties because of their small picture size. It is noteworthy that the mean query length climbs over 33 terms per question for the upscale options (light, medium/beta) as well as the remastered version in comparison to 26 terms for basic mode inquiries (interestingly, this does not generalize to max upscale). Additionally, we manually categorized the 150 most common phrases into the following categories: style, content, and quality. It is noteworthy that the percentage of style phrases dramatically rises for the upmarket and polished searches.

The average number of iterations required to create a query is shown in Fig. 7(f). If the same user runs the same or an expanded query more than once in a 30-min period, we identify that as an iteration. 9.84 % of all queries (257 k, single) are only performed once. Multiple iterations are used to increase the maturity of the query. In draught mode queries, there are roughly 4.4 stages. There are around 7.8 steps in total for queries that can be upscaled. After 5.1 draught steps, they are upscaled into three separate variations (light, medium/beta, and max). Remastered queries go through around 6.5 iterations. They only have 2.0 queries in draught mode, but 2.8 remastering steps and 1.5 upscale steps.

This demonstrates that users typically build their queries through several rounds by choosing the best versions or include extra terms (particularly style terms), rather than coming up with the ideal query right away. We created several suggested workflows that we will describe in the following part based on these observations and real-world experience.

3. Results and discussion

Here, we highlight a few of the practical tasks that the three technology platforms for artificial intelligence give architects access to. These tasks are based on the findings of our analysis as well as the knowledge we have obtained through carrying out numerous trials on their interfaces. As we've seen, we frequently iterate rather than running a single query and expecting a perfect answer. If one enters the procedure without having a clear understanding of the necessary processes, it is simple to encounter dead ends and accounts quickly run out of balances.

The earliest stages of design are the best times to use text-to-image generation in architecture. An architect could swiftly create a visual representation of their design concepts from a textual description using a text-to-image generating tool. This could help the architect quickly explore a wide range of possibilities and iterate on different design approaches. Visualizing building layouts and plans is another potential application of text-to-image generation in architecture. An architect might easily create 2D or 3D visualizations of a building's layout using a text-to-image creation tool based on a written description or set of floor plans. This might help make the building's design more approachable for customers and builders. creation of accurate photographs of buildings or other structures for use in presentations or marketing materials, such as.

- Develop interactive, immersive experiences that let consumers investigate and assess many design possibilities.
- Automatically produce intricate 2D drawings.
- Create renderings of unfinished buildings or structures to help clients and architects see the finished project.

The production of 3D models is another potential use for text to picture generation in architectural design. The Text-to-2D Image Generation model can only be used as a source of ideas for concept designs at this time. However, current studies suggest that text-to-3D models will soon be feasible [1,32]. Currently, creating 3D models of buildings is a time-consuming and arduous process that needs specialized tools and a high level of technical proficiency. Using text to image creation, architects could just use everyday language to describe the ideal building, and the system would create a 3D model for them. In addition to saving time and effort, this would also make it simple for architects to experiment with different design ideas. The key takeaway from this study is that "scenarios beyond the model and platform's intended scope (e.g., floor plans) necessitate the development of a new model specifically trained on the relevant dataset" as a significant 'challenge'. As a result, there arises a necessity to delve into more detailed supplementary training in the paper, along with its potential applications.

According to the paper, the Txt2Img generation approach can serve as a tool for augmenting creativity and visual representation within the realm of architecture. However, the current technology generally falls short of producing the precise outcomes desired in the initial attempt. Consequently, adjustments such as Retraining and Inpainting, facilitated by option functions within each platform, become imperative. In essence, further refinement is essential to elevate the accuracy and caliber of the generated images. The investigation confirms the execution of a quantitative analysis of prompt engineering's contextual aspects, centered around Midjourney. However, the essence of this examination can be summarized as "finding the optimal prompt necessitates multiple trials and errors to achieve the desired image." Nonetheless, a discernible gap exists in terms of proposing solutions to this challenge.

3.1. Potentials for interior design

Interior architecture could be revolutionized by text-to-image conversion. Interior architects may rapidly and simply develop visualizations of their designs using the capacity to produce realistic visuals from text descriptions, enabling more effective client communication and stakeholder engagement.

One of the main benefits of text-to-image generation in interior architecture is the speedy creation of high-quality graphics. Interior architects can simply explain their concept in language and have the computer generate a realistic image rather than spending hours or even days constructing elaborate 3D models or computer-generated graphics. This not only saves time and effort, but also enables greater experimentation and flexibility because designers can quickly create several iterations of their designs to explore different choices.

In addition to expediting the design process, text-to-image production offers the potential to enhance the clarity and accuracy of interior design visualizations. Instead of depending on a rough 3D model or a sketch, interior architects may make sure that the generated image faithfully captures their vision by providing a written description of the design. This can help in avoiding misunderstandings and miscommunications as well as in spotting possible difficulties or design flaws before they become significant roadblocks.

3.1.1. Comparing the different interior design models

We will first examine the models' performance when used only on their own. The scenario will be from the realm of interior design. We begin with a straightforward query that doesn't make use of any platform-specific parameters or special procedures.

On all three we attempt to provide a high-quality representation of a space described as "cozy living room, wood paneling, television, large sofa, natural light, lived in, realistic, full view".

As seen in Fig. 8 (a)-(b), Midjourney begins with a number of findings that are stylized or have odd perspectives. The central sofa in

Fig. 8. Minimum workflow for the provided query for Midjourney (a) through (d), DALL-E 2 (e) through (h), and Stable Diffusion (i) through (l) [25,28,29].

the chosen interior design in (c) is still an illogical form in the image's Centre despite the initial upscale's significant improvements in material quality and overall detail. We can activate the "remaster" stage, which is displayed upscaled in (d), if the standard image output falls short of meeting the required standards for cohesiveness and realism. However, even this final output has some minor perspective issues that are difficult to correct without manual image editing.

The DALL-E 2's initial results in Fig. 8 (e) demonstrate that it has trouble accurately responding to the query's "realistic" phrase. When one of the two realistic variations is chosen, the subsequent step in (f) for variant formation produces more beneficial outcomes. In order to create in paint versions, we mask out specific sections of the image in (g) to correct perspective or coherence issues. The last option is displayed in (h). It's noteworthy that even areas that haven't been painted respond appropriately to the scene's preexisting illumination. Although the outcome is of decent quality, the web interface does not allow for further upscales.

In Fig. 8 (i), Stable Diffusion starts off with significantly better results than its two rivals, producing pictures that successfully include the "realistic" and "lived in" aspects of the query. These spaces don't appear to be computer-generated projections but rather like actual rooms. However, each picture is close-up to provide the viewer a good interior look. For this reason, in addition to using inpainting in (j) to correct mistakes, we also include more canvas space for outpainting. As a result, the outpointed sections have some really illogical versions. The option in (l) was chosen as the best after some additional iterations.

Because all of the initial versions are very realistic, stable diffusion performs the best overall in this situation. Additionally, it exhibits no significant flaws in the later stages. Only the last outpainting stage, where it might be necessary to manually smooth the transitions into the outpointed portions, requires attention. Midway produced results of comparable quality, but provided a less strong starting point for free design ideation.

Additionally, text-to-image conversion holds the promise of democratising interior design. By offering a more user-friendly and accessible method for visualizing designs, text-to-image generation can make the design process more accessible to a wider range of people, including those who might lack the technical know-how to create intricate 3D models or challenging computer-generated images. As a result, the interior architecture community may become more diverse and welcoming, and new ideas and creativity may be encouraged.

Text-to-image generation in interior architecture has a wide range of possible advantages, but there are also a number of challenges and limitations that must be taken into account. For text-to-image creation algorithms to generate accurate and realistic images, they need high-quality training data. This is among the biggest challenges. It is challenging and time-consuming to create a big and diverse dataset of interior design photos annotated with accompanying text descriptions, and it is uncertain how effectively current algorithms can handle the unique difficulties of interior architecture.

3.2. Potentials for Exterior Design

One must become familiar with the terminology and parameters each model is trained to function inside and the right style of image generation in order to utilize each model's capabilities. The fact that the keyword "digital art" can considerably improve many prompts is one of the first suggestions DALL-E gives new users, which is closely tied to the data it was trained on.

When it comes to Midjourney, refinement begins with a process of adding and removing certain terms from the prompt to get as close as we can to the intended style. These expressions can be highly tangled. It typically helps to mention the modelling software types (like "octane render" or "cinema3D″) or even quality indicators (like "top 10 on artstation") that would produce the required kind of image in the prompt. The use of weighting words, picture references, and parameters, which further parameterize the prompt system, provides a far more targeted method of influencing requests.

3.2.1. Exterior Design - combining the models

Iteration across image variants for DALL-E and Stable Diffusion is typically a quicker, more accurate procedure since iterations call for direct intervention on the image data in the form of in-/outpainting. Once we are aware of these resources' advantages, we can use them to combine them into a more adaptable process. As Midjourney thrives at free-form ideations, we will begin the next ideation workflow with a Midjourney prompt to generate a desired scene, which we will then polish using in-/outpainting in DALL-E and Stable Diffusion to achieve a specific outcome that is inspired by the initial Midjourney result. Fig. 9 is a flowchart of the process, which we

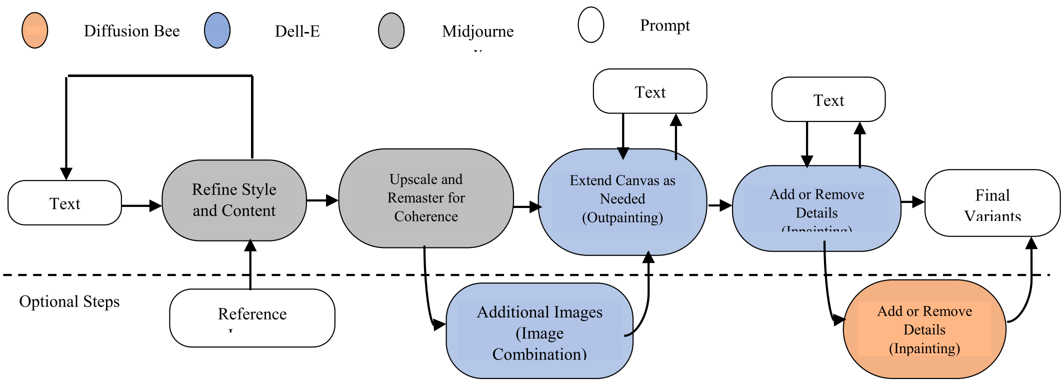


Fig. 9. The combination process that is suggested between Midjourney, DALL-E, and (optionally) Stable Diffusion.

shall explain along with the outcomes in Fig. 10. Additionally, text-to-image conversion might not correctly depict abstract or complicated design principles. While the technology can create accurate visual representations of actual architectural features like roads and buildings, it can have trouble with more abstract or conceptual ideas. This might limit how fully designers can use this technology to explore and convey their creative ideas (see Fig. 2).

The embryonic stages of an idea generated using Midjourney are depicted in Fig. 10 (a)–(c). “Single-family home with garden, full exterior view, modern architecture, photo, sunlight" was the search term that produced this specific outcome. Before this result was chosen, remastered, and then upscaled, several keyword arrangements and weights on various phrases were explored. Although that method wasn’t used here, it is also feasible to start with one or more reference photos.

The finished product was then transferred to the DALL-E, where it was outpointed in Fig. 10 (d) to provide a wider viewing angle and later inpainted in Fig. 10 (e)–(f) to remove extraneous elements such the cables hanging in the air and the variously colored windows.

We transferred image (f) to Stable Diffusion after failing to successfully add a paved pathway from the sideway to the entrance through inpainting in DALL-E. As the walkway would frequently be overrun by vegetation, we adjusted the query to remove the garden reference and add "paved" early in the prompt (earlier keywords are weighted higher). We then removed the walkway and replaced it with inpainting. In response, the prompt "single-family home, paved between sidewalk and door, full exterior view, modern architecture, photo, sunlight" was generated, with the outcome shown in image (g).

Now, if we picture a scenario where the outcome of this procedure is used to show to a potential client, demands for modifications must be met. Consider the scenario where a client requests a second storey or another type of roof component. This is simple to do with inpainting by deleting the roof and a portion of the sky and using a slightly modified prompt that specifies the new image’s style. Fig. 10 (h) and (i) depict the outcome of the Stable Diffusion query "single-family home, two stories, clear blue sky, curved roof, full exterior view, modern architecture, photo, sunlight" following the removal of the roof and the center portion of the sky and the extension of the canvas upward to make more space for the roof elements. Note that it can be challenging to avoid some latent effects, such as the tree branches that extend into the picture.

Despite these drawbacks, text-to-image generation can considerably improve urban design by giving designers a more effective and simple way to conceptualize and visualize their ideas. Text-to-image generation has the potential to revolutionize the way urban design is approached and carried out due to its capacity to quickly and easily generate a large number of design options, improve collaboration and communication within design teams, and raise the accuracy and realism of design representations.

3.3. Common limitations

While examining Swissness and Alpine architecture in a controlled experiment, we were able to systematically observe the cause and effect of different interventions. However, there are some limitations to this approach. The primary limitation is the size and quality of the data set, which we gathered from open sources on the web. Future research could increase selectivity and durability by expanding the size and quality of the data set. Also, the limited number of human subjects involved in the experiment hinders the generalizability of the research. Future studies could expand the size and diversity of the subjects to address this limitation. Lastly, it’s important to consider limitations related to the AI.

Working with current AI models entails a certain amount of trial and error. Typically, they don’t produce the required outcomes accurately on the first try. Instead, they provide a number of options for each question and ask the user to choose the best one or alter their query language until they get a satisfactory result. Although it may not be simple, assigning a group of architecture students the duty of creating a design with such vague parameters would likewise lead to several variations and take a lot longer.

However, there are instances in which the variant tree that users explore completely fails, and a desirable outcome cannot be attained without a lot of trial and error, retraining, or inpainting.

Fig. 11 displays a few of the typical failure scenarios that architect could run into while employing these models. The response to a floor plan query is seen in Case (a). It does a good job at copying the style of using thick lines for things and bold lines for walls, but it eventually loses all sense. This is because AI art tools only copy the aesthetic; they do not comprehend the semantics of the lines in a floor design. Case (b) illustrates the outcome of a query made up of several technical phrases that are both unclear in and of themselves. The "bow window" architectural feature was changed into a window with a bow above it. A distinct gable piece with a clock beneath it served as the "clock gable"’s representation. A query for a landscape design with five buildings is shown in Case (c), which highlights the widespread issue that these AI systems are just lousy at counting and sophisticated spatial arrangements beyond foreground and background.

4. Conclusions

The scope of objective measures in AI research is growing as a result of the development of GAN algorithms [33]. The most complex structures our society has ever produced can now be manufactured because to Artificial Intelligence Assisted Design (AIAD), which has been made feasible by the exponential rise of hardware and software technology.

All of the use cases that are currently outside the scope of these models and platforms need for a semantic comprehension of the visual content. A floor plan consists of more than just a set of lines. These lines depict semantic and contextual information, such as the fact that a door is truly needed to enter one room from a related room that is separated from it by a wall. However, since Building Information Models (BIM) currently exist and offer this semantic data, it is only a matter of time before new models emerge that have been trained particularly on these datasets and are likely to be able to meet all of the requirements listed in Table 1.

The 2.6 million inquiries with architectural context that we detected reflect a substantial acceptance of AI models in architecture, as

Fig. 10. Refinement and variation generation for a walkway (g) and a second storey (h)–(i) in Midjourney (a) through (c), DALL-E 2 (d) through (f), and stable diffusion (g) [25,28,29].

(A) Architectural floor plan of a modern single-family home, ground floor.

(B) Red brick building façade with a Dutch clock gable and a bow window.

(C) A drawing of a map of five a buildings arranged around a park, realistic, detailed, top-down, site plan, digital rendering.

Fig. 11. Example failure cases [31].

do the numerous use cases and outcomes presented in this paper. In order to present a broad overview of the situation and talk about how they might be used in practise, we have examined AI art creation tools from a variety of angles in this paper. To get things going, we've chosen use cases for architectural design that can already be handled by picture generating models as well as those that are probably going to become available shortly. The distinctions between three of the leading candidates right now have been emphasized. In order to understand how users really query these architectural outcome models and how many queries it appears to have taken them to achieve a satisfying result, we conducted a quantitative analysis of millions of questions.

The widespread use of text to image generation in architectural design faces a number of obstacles [34]. First of all, there is still more work to be done before the technology is ready for public usage. The early stages of its evolution are still present. Second, there is a chance that the generated images won't be accurate or realistic enough to be used in the design process because their quality is not yet on par with that of professional software. Thirdly, since spoken language isn't always precise, there's a chance the algorithm will misunderstand the architect's description and produce the wrong image.

Despite these challenges, text-to-image generation in architectural design has significant potential benefits. The ability for architects to quickly and efficiently produce 3D models, photorealistic photographs, animations, and virtual tours has the potential to revolutionize the design process. In the near future, text-to-image production is probably going to become a crucial tool for architects, allowing them to quickly experiment with various design concepts and produce visual representations of their designs.

Finally, we made an effort to condense these concepts into workflow examples that readers may apply to their own usage of these various platforms. Overall, text-to-image is a robust and adaptable AI technology that has the potential to greatly enhance the effectiveness and efficiency of the engineering design process. For architects and other creative professionals, it is an exciting and useful tool due to its user-friendly interface and capacity to produce a broad variety of images. In the future, instead of typing code, we will be able to describe a software and its expected outputs in natural language. Deep learning models, such as Codex, BLOOM [35] or other large “foundation models” [36], will then generate executable software code based on the human’s spoken or written input prompts.

The paper explores the potential use of the Txt2Img technique for architectural design visualization. However, the focus is mainly on testing the generative method and technical aspects, rather than specific illustrations or outcomes. To make it more practical, it would be helpful to include a basic plan for improvement with analysis and examples of application. The research also explores combining machine and human intelligence for future design workflows. The experiment used NLP techniques to analyze human interpretation of Generative Adversarial Networks generated images, providing a more comprehensive measurement of GANs in architecture. To extend the research, a crowdsourced human evaluation could be analyzed using NLP methods. The combination of generative and analytical capacities of machines, along with human perceptual intelligence, could create new approaches to designing future architectures and cities, where the line between human and artificial agency is blurred.

The topic of "Interaction between Creativity and AI Generation Model" is a highly debated qualitative research subject. This paper aims to clarify the scope of creativity within the confines of the study and categorize the subject matter. Overall, this is an intriguing paper that explores new ways of building design and engineering.

Authors’ Contributions

The Author, Nervana O. Hanafy: Ideas, Design of Methodology and Writing-original Draft preparation, Development, Editing the Final manuscript before summation.

Funding

This study had no funding from any resources.

Availability of data and materials

All data and materials are available upon request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Declaration of competing interest

The author declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

Not applicable.

References

[1] Z. Liu, Y. Wang, X. Qi, C.-W. Fu, Towards implicit text- guided 3D shape generation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 17896–17906, 2022. Retrieved from, https://github.com/liuzhengzhe/Towards-Implicit.
[2] N. Jung, et al., Automated classification of building information modeling (BIM) case studies by Bim use based on natural language processing (NLP) and unsupervised learning, Adv. Eng. Inf. 41 (August 2019), 100917, 2019.
[3] M. Nik-Bakht, et al., Project collective mind: unlocking project discussion networks, Autom. ConStruct. 84 (December 2017) 50–69.
[4] X. Lin, et al., Stakeholders’ influence strategies on social responsibility implementation in construction projects, J. Clean. Prod. 235 (20 October 2019) 348–358.
[5] R.A. Niemeijer, et al., Freedom through constraints: user-oriented architectural design, Adv. Eng. Inf. 28 (1) (January 2014) 28–36, 2014.
[6] Y. Ding, J. Ma, X. Luo, Applications of natural language processing in construction, Autom. ConStruct. 136 (2022), 104169, https://doi.org/10.1016/j.autcon.2022.104169.
[7] J. Zhang, et al., Integrating semantic NLP and logic reasoning into a unified system for fully-automated code checking, Autom. ConStruct. 73 (January 2017) 45–57.
[8] I.T. Afolabi, et al., Identifying major civil engineering research influencers and topics using social network analysis, in: Cogent Engineering, 2020.

[9] M. Al Qady, et al., Automatic classification of project documents on the basis of text content, J. Comput. Civ. Eng. 29 (3) (2015).
[10] T.B. Brown, et al., Language models are few-shot learners, in: Advances in Neural Information Processing Systems, 2020.
[11] P. Martinez, M. Al-Hussein, R. Ahmad, A scientometric analysis and critical review of computer vision applications for construction, Autom. ConStruct. 107 (2019), 102947, https://doi.org/10.1016/j.autcon.2019.102947.
[12] W. Fang, H. Luo, S. Xu, P.E. Love, Z. Lu, C. Ye, Automated text classification of near-misses from safety reports: an improved deep learning approach, Adv. Eng. Inf. 44 (2020), 101060.
[13] H. Yan, N. Yang, Y. Peng, Y. Ren, Data mining in the construction industry: present status, opportunities, and future trends, Autom. ConStruct. 119 (2020), 103331.
[14] R. Romero-Silva, S. De Leeuw, Learning from the past to shape the future: a comprehensive text mining analysis of OR/MS reviews, Omega 100 (2021), 102388, https://doi.org/10.1016/j.omega.2020.102388.
[15] J. Ho, T. Salimans, A.A. Gritsenko, W. Chan, M. Norouzi, D.J. Fleet, Video diffusion models, in: ICLR Workshop on Deep Generative Models for Highly Structured Data, 2022.
[16] X. Zeng, A. Vahdat, F. Williams, Z. Gojcic, O. Litany, S. Fidler, K. Kreis, LION: Latent Point Diffusion Models for 3d Shape Generation, 2022.
[17] X. He, L. Deng, Deep learning for image-to-text generation: a technical overview, IEEE Signal Process. Mag. 34 (6) (2017) 109–116.
[18] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, Y. Bengio, Generative adversarial nets, Adv. Neural Inf. Process. Syst. 3 (January) (2014) 2672–2680, 10.3156/jsoft.29.5_177_2.
[19] R. Anadol, Space in the mind of a machine: immersive narratives, Architect. Des 92 (3) (2022) 28–37, https://doi.org/10.1002/ad.2810.
[20] WDCH Dreams - Refik Anadol, Retrieved August 21, 2022, from, https://refikanadol.com/works/wdch-dreams/, 2019.
[21] P. Dhariwal, A. Nichol, Diffusion models beat GANs on image synthesis, 35th Conference on Neural Information Processing Systems 1–15 (2021).
[22] Erdem Yildirim, Text-to-Image generation A.I, Architecture, Art and Architecture: Theory, Practice and Experience Edition (2022) 97–120. Chapter1, Publisher 8, Livre de Lyon.
[23] A. Radford, J.W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, et al., Learning Transferable Visual Models from Natural Language Supervision, ICML, 2021, pp. 8748–8763.
[24] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, I. Sutskever, Zero-shot Text-To-Image Generation, ICML, 2021, pp. 8821–8831.
[25] DALL·E 2. n.d. DALL·E 2. com. http://adityaramesh.com/posts/dalle2/dalle2.html.
[26] Vivian Liu, Lydia B. Chilton, Design guidelines for prompt engineering text-to-image generative models, Article 384, in: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22), Association for Computing Machinery, New York, NY, 2022, p. 23, https://doi.org/10.1145/3491102.3501825. pages.
[27] Piero Molino, Ré Christopher, Declarative machine learning systems, Commun. ACM 65 (dec 2021) (2021) 42–49, https://doi.org/10.1145/3475167, 1.
[28] n.d. Midjourney, Midjourney.com. https://www.midjourney.com.
[29] DiffusionBee. n.d. DiffusionBee. com. https://diffusionbee.com/.
[30] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, High- resolution image synthesis with latent di usion models, CVPR, 2022, pp. 10684–10695.
[31] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, NIPS (News Physiol. Sci.) 26 (2013).
[32] I. Koh, Architectural plasticity: the aesthetics of neural sampling, Architect. Des 92 (3) (2020) 86–93.
[33] C. Wang, M. Chai, M. He, D. Chen, J. Liao, CLIP-NeRF: text-and-image driven manipulation of neural radiance fields, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022) 3835–3844, https://doi.org/10.1109/CVPR52688.2022.00381.
[34] B. Kawar, S. Zada, O. Lang, O. Tov, H. Chang, T. Dekel, I. Mosseri, M. Irani, Imagic: text-based real image editing with di usion models, in: arXiv Preprint arXiv:2210.09276, 2022.
[35] BigScience Initiative, Introducing the World's Largest Open Multilingual Language Model: BLOOM, 2022 bigscience.huggingface.co/blog/bloom.
[36] Rishi Bommasani, A. Drew, Hudson, ehsan adeli, russ altman, simran arora, and percy liang, in: On the Opportunities and Risks of Foundation Models, 2021 arXiv:2108.07258 [cs.LG].