AI in Civil Engineering

ORIGINAL ARTICLE Open Access

AI art in architecture

Joern Ploennigs[1*] and Markus Berger[1]

Abstract

Recent diffusion-based AI art platforms can create impressive images from simple text descriptions. This makes them powerful tools for concept design in any discipline that requires creativity in visual design tasks. This is also true for early stages of architectural design with multiple stages of ideation, sketching and modelling. In this paper, we investigate how applicable diffusion-based models already are to these tasks. We research the applicability of the platforms Midjourney, DALL·E 2 and Stable Diffusion to a series of common use cases in architectural design to determine which are already solvable or might soon be. Our novel contributions are: (i) a comparison of the capabilities of public AI art platforms; (ii) a specification of the requirements for AI art platforms in supporting common use cases in civil engineering and architecture; (iii) an analysis of 85 million Midjourney queries with Natural Language Processing (NLP) methods to extract common usage patterns. From this we derived (iv) a workflow for creating images for interior designs and (v) a workflow for creating views for exterior design that combines the strengths of the individual platforms.

Keywords Image generation, Diffusion models, Natural language processing, Architecture

1 Introduction

1.1 Motivation

Recent versions of AI art generation tools are reaching levels of output quality that allow them to support architects and designers in parts of their daily work. This gained them the attention of several architects across the globe, reflected also in a special edition of the AEC magazine.[1]

Beyond this public discussion, we want to take a deeper look into the current capabilities of this technology and analyse qualitatively as well as quantitatively what benefits it offers to architects now and in the future. In this paper we therefore review the technology behind the three leading commercial AI art platforms and evaluate what use cases they currently support in architecture. We investigate how users are already using these tools by analysing more than 85 million public queries from one of these platforms with Natural Language Processing (NLP) workflows. Finally, we present a collection of case studies in which we apply the practical experience we collected in working with these systems.

The novel contributions of this paper are thus:

- a comparison of the technology of three leading AI art platforms;
- an analysis of how well current AI art tools can handle different use cases;
- a mapping for what specific design tasks each platform supports;
- an NLP analysis of how these platforms are used for architecture today;
- a collection of practical workflows for specific architectural design tasks.

[1] https://aecmag.com/technology/ai-special-edition-of-aec-magazine/.

*Correspondence:
Joern Ploennigs
Joern.Ploennigs@uni-rostock.de
[1] AI for Sustainable Construction, University of Rostock, Rostock, Germany

1.2 State of the art in generative methods

The potential benefits of AI art platforms for creative work is hard to overstate. These AI art platforms are all using *generative machine learning models*, specifically text-to-image generative models. Despite their specialization on generating images, many are based on the natural language model GPT-3. GPT-3 is trained to generate text that completes a textual input query Brown et al. (2020). More precisely, it predicts the next possible combinations of words to an input text. The specific *Image GPT-3* models used by AI art platforms are trained to instead predict the next cluster of pixels, called patch, in an image for a given input text.

The most recent generation of generative models combines natural language and so-called *diffusion models*. The idea for diffusion models was first proposed in Sohl-Dickstein et al. (2015), in which (structured) image information is slowly destroyed through a *forward diffusion process* that introduces noise into the image data and then generated anew through a *reversed diffusion process*. This reverse process generates completely new image data, as the original information was fully destroyed by noise. This approach was constantly improved over the years with a strong focus on optimizing the underlying neural network architectures Ho et al. (2020), resulting in several variants, like OpenAI's GLIDE model Nichol et al. (2022). It consists of an encoder that creates a text encoding based on the user prompt, a model implementing the diffusion based on this text encoding, as well as an upsampler that upscales and denoises the result. Current diffusion models often implement the process of text encoding and the association of those text encodings with image parts with the CLIP (Contrastive Language Image Pre-training) architecture presented by Radford et al. (2021) and used by DALL·E 1.

One of the currently most advanced incarnation of the technology is DALL·E 2 (from here on out simply referred to as DALL·E), based on the unClip method developed by Ramesh et al. (2022). It uses one image encoder for both text and images into a diffusion-based joint representation space (the *prior*). The image generation is done by a similarly trained *decoder*, which translates the prior's encoding back into an image. Another main platform in the field is the open-source model Stable Diffusion, which is also based on the CLIP text encoder. The third contender, Midjourney, does not published their models, but it is assumed that it is using a similar structure.

This kind of diffusion model architecture can solve various image generation tasks. Is a completely new image generated from a user-written text prompt then a *text-to-image* model (txt2img) is used. Is an existing image modified based on a text prompt then *Image-to-image* models (img2img) are used. They either change the style or arrangement of the image based on the text prompt. If a certain part of the original image was deleted, the model can replace it with entirely new content based on the prompt. This approach is called *inpainting* by Lugmayr et al. (2022). A similar approach is *outpainting*, also called *uncropping* by Saharia et al. (2022), which adds additional content outside the image. If a user is requesting changes to the original image without manually deleting or masking out parts, then this is called *image editing*. Image editing is not yet available in commercial AI art platforms, but there is recent work in single-image editing through text prompt, for example by Kawar et al. (2022). Is another diffusion step applied to add more details to the image in a higher resolution then it is *upscaled*, a method also called *super-resolution* by Saharia et al. (2022). Often platforms offer multiple or all these methods, with configurable weights between the individual images and words.

Assessing the quality of all these models and architectures systematically is difficult. Attempts at quantitative evaluation are being made such as Borji (2022). However, in such cases it is difficult to evaluate subjective metrics like style, i. e., whether an oil painting style looks better than a photorealistic one.

As for research into use cases from architecture, Seneviratne et al. (2022) describe a systematic grammar for using DALL·E for the purpose of generating images in the context of urban planning. They open-sourced 11.000 images generated with Stable Diffusion and 1.000 created with DALL·E by that grammar. They found that, though, many realistic images could be generated, the model has weaknesses in creating real-world scenes with a high level of detail. But, these models advance quickly and DALL·E 2 outperforms DALL·E 1 significantly.

Recent progress was also made in generating video (Ho et al., 2022), 3D models via point clouds (Luo and Hu, 2021; Zhou et al., 2021; Zeng et al., 2022), and even 3D animation data (Tevet et al., 2022). While this is beyond the scope of this paper, especially the generation of 3D models will be revolutionary for architecture. The 3D workflows would be similar to the image-based workflows we present.

2 Model Architectures and Interfaces

Although the concept of generative art has been a research area for years, it only entered public perception with the advent of publicly available diffusion model platforms like DALL·E, Midjourney or Stable Diffusion, which combine txt2img, img2img, inpainting and upscaling into easy-to-use workflows.

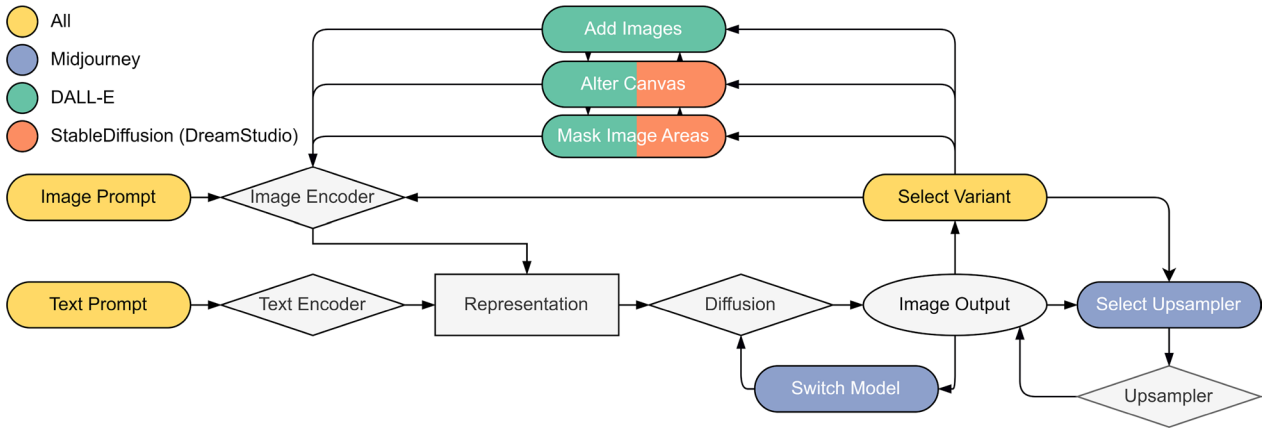


Fig. 1 Model architecture and image generation process in different models. Grey elements show the AI workflow, coloured elements the user interaction

These models are not only competing on the technical level, but also in terms of user experience. Midjourney[2] directly interacts with its community by sharing queries across public (or private) channels in the Discord messaging app. They do not provide a dedicated user interface, but simply return the generated images as a chat response to the query. Direct interactions are possible with attached links that usually result in new queries. In contrast, DALL·E 2[3] is only accessible by individuals through a dedicated web-based user interface with authoring and editing tools. It provides a simply query interface without additional query parameters. Stable Diffusion on the other hand is released as open source, which fosters a plethora of community-created tools that are usually used more than the official web-interface[4].

The internal model architecture as well as the interface paradigm influences how these models can be utilized. In Fig. 1 we illustrate some of the similarities and differences between DALL·E 2, Stable Diffusion (v2.1), and Midjourney (v4). The core workflow of foundational models in grey are similar across technologies as stated before. Differences lie in the workflows that they provide, which often results from their different interface approaches.

It is apparent in Fig. 1 that while the core workflow in grey is similar, the ways to refine their results do vastly differ. Only Midjourney offers successive upscaling of resolution and does so with multiple different sizes. Thus, Midjourney's workflow focuses on generating and comparing different image variants and then upsampling the best results in multiple ways, with limited possibilities to remix the original text prompt throughout.

In contrast, both DALL·E and Stable Diffusion allow direct editing of uploaded images or previous results. They do not provide traditional image editing tools like drawing, filling, layering, or stamping. Instead, all image editing must be done by img2img-based operations like erasing sections (inpainting) or extending the canvas (outpainting). All networks have ways to create images of different sizes and aspect ratios, either by specifying the size in the query or by altering it later through outpainting.

Notably, image generation only takes a few tens of seconds in all models, making it fast enough to use in creative sessions alone or with clients. All three models also allow importing external images in some capacity. Therefore, they can easily be combined into composite workflows, as shown in Sect. 5.2.

One aspect not included in Fig. 1 is the training data. There is little information on the training data used for DALL·E and Midjourney. However, Stable Diffusion was trained on the LAION-5B dataset (Rombach et al., 2022), which is based on image and text data scraped from the web. Similar internet datasets are very likely the source for Midjourney and DALL·E. However, it is evident that either biased by the training data or the training process, these models have developed very different image styles. DALL·E and Stable Diffusion are good in generating both drawn images as well as photorealistic outputs. Midjourney tends towards a more artistic style, especially with earlier model versions. But, with carefully chosen keywords most styles can now be targeted by all models.

[2] https://midjourney.com.

[3] https://openai.com/dall-e-2.

[4] https://beta.dreamstudio.ai.

Table 1 Top part—comparison of platforms with their supported features; Lower part—Mapping of architectural use cases to features

Model	txt2img	img2img	In-/Out-paint.	Editing	Upscaling	Semantics
DALL·E 2	♦	♦	♦	◇	◇	◇
Midjourney v4	♦	▲	◇	◇	♦	◇
Stable Diffusion v2.1	♦	♦	▲	◇	▲	◇
Use Case						
Ideation	♦/♦	▼/♦	◇/▼	◇/◇	◇/▲	◇/◇
Sketches	▲/▲	▼/▲	▼/▼	◇/◇	◇/▲	▼/◇
Collages	◇/◇	▼/◇	▲/▲	♦/◇	◇/◇	◇/◇
Image Combination	◇/◇	▲/◇	♦/♦	▼/◇	▼/◇	◇/◇
Build Variants	◇/◇	▲/▼	♦/▲	♦ ◇	◇/◇	▲/◇
Style Variants	◇/◇	♦/▲	◇/◇	◇/◇	▼/▼	◇/◇
Construction Plans	▲/◇	◇/◇	◇/◇	▲ ◇	▲/◇	♦/◇
Exterior Design	▲/▲	▼/▼	▲/▲	▲/◇	▼/▼	▼/◇
Interior Design	▲/▲	▼/▼	▲/▲	▲/◇	▼/▼	▼/◇
Creating Textures	▲/▲	◇/▼	◇/◇	▼/◇	♦/▲	◇/◇

♦ full, ▲ limited, ▼ bad, or ◇ no support; ♦ high, ▲ some, ▼ low, or ◇ no importance; versus (/) how well it works: ♦ well, ▲ somewhat, ▼ a little, ◇ not at all

3 Architectural use cases

Given the discussed differences of the platforms, they vary in the architectural use cases that they support. To analyse this, we collected a series of use cases where architects and planners normally create or edit images. A direct comparison between the platforms and the use cases is often complicated, because the style and quality of the results heavily depends on the input prompts. We identified that the more differentiating factor is, whether a platform supports a specific technical feature that is required to realize the use case.

Therefore, we evaluated for each use cases what features they require and how qualitatively well they are supported by each platform. In addition to the image operations explained in the previous chapter, we also consider support for architectural semantics, i. e., structured knowledge that goes beyond common image training datasets. Table 1 shows the results for the use cases that we discuss below:

- *Ideation:* Developing ideas by randomly generating images for inspirations. This is what txt2img models are made for and it works splendidly. Additional image prompts can add style and object references.
- *Sketches:* Drawing architectural sketches with specific target style and items. This works well for common examples in the training data, but, less so with specific requests.
- *Collages:* Combining and filling existing images with life by adding people and objects. This can be done through inpainting for individual items, but not through generic requests like: "Add many people".
- *Image combination:* Taking multiple existing image elements (for example multiple buildings), arranging them on a canvas and then creating a coherent composite image.
- *Build variants:* Taking an existing sketch or picture and generating versions in which certain elements are altered (like adding a garage). This works well through inpainting.
- *Style variants:* Taking an existing image and transforming its style (e. g., a sketch to photorealistic art deco) without changing content. This works well with certain models.
- *Construction plans:* Creating detailed layout plans to establish spatial relations. This rarely works as the models do not understand the semantics of line styles, areas, etc.
- *Exterior design:* Finding style and feeling for a building and the surrounding area/landscape. This works well for common scenarios.
- *Interior design:* Finding a style or feeling for an interior space. This also works well in many scenarios.
- *Creating txtures:* Creating tiled patterns to serve as surface materials for 2D or 3D models. This is currently a unique feature of Midjourney.

Even though Stable Diffusion seems to support fewer features than DALL·E 2, it's main advantage can not be overstated: It is possible to run it locally and train it on

own data to introduce e. g. new architectural concepts. Together with the high quality of Stable Diffusions outputs, this makes it the most potent of the three models to be specialized on architectural drawings.

4 Analysis of architectural queries

We also explored how people use these AI art platforms in practice. We analysed about 85 million user queries that we collected over one year since Jan. 30th, 2022 from Midjourney. It is the only platform for which many user queries are publicly visible. Midjourney uses the Discord messaging app as its main interface, which allowed us to monitor the public channels for queries that we consider to be of an architectural nature. We selected queries containing either the word "architect", "interior" or "exterior" design or one of 38 architectural keywords like "building", "facade", or "construction" (listed in Fig. 2 (b) and[5]). We identified these keywords by selecting only those from architectural glossaries[6, 7] that co-occurred in at least 10 % of all cases with "architect", "interior" or "exterior" in the queries. We also added to the list of keywords the names of 941 famous architects from Wikipedia[8] as we noted that several queries refer to their style by naming them. By applying these filters, we identified 5.7 million queries (6.7 %) with potential architectural intent including 2.2 million queries (2.6 %) explicitly containing "architect", "interior" or "exterior" design.

In the next steps we filter out stopwords (e. g. "a", "and", "the") and build a Word2Vec model (Mikolov et al., 2013) from these queries to get a model of the occurrence and co-occurrence of terms in these queries. For understanding the results, it is important to know that most Midjourney users do not formulate full sentences, but a prompt is rather a collection of terms that refer to the content, style, or render quality of the targeted image.

Figure 2 visualizes the main results of our analysis. Figure 2(a) shows the most frequent terms used in the filtered queries. The colour blue represents the frequency across all 85 million queries, red is the frequency within the filtered 5.7 million queries and green within 2.2 million queries explicitly containing "architect", "interior" or "exterior" design. Note that red, green, and blue are mutably inclusive and overlap. The top 10 of words has a similar frequency across all three classes. Many of those refer to Midjourney style commands like "detailed", "realistic", or "cinematic". However, some terms like "black", "creative" or "full" have high overall frequency, but, low frequency in architectural context. Other terms like "architecture", "interior", "house" do only occur exclusively within our filtered results as they are part of our keyword list.

Figure 2(b) shows our extended keyword list and their respective frequency. As we filter on these keywords, their total frequency is identical with the filtered one and we do not display it. It is of note that "architecture" and "interior" keywords are the most and third frequent keyword. It is notable that "interior" is six times more popular than "exterior" design as keyword, but, this simply may be that users refer to it implicitely trough "architecture".

Figure 2(c) shows the frequency of famous architects that we extracted from 430.333 queries that referred to at least one of them. Zaha Hadid is by far the most frequently queried architect, given her well recognizable parametric style that is popular in the community of people experimenting with AI tools. Michelangelo and William Morris are second and third. The low red bar shows that they are usually not used in architectural context but for their art contributions. Adrian Smith is also often used in other context probably referring to the musician. The architects Frank Lloyd Wright, Tadao Ando, Frank Gehry, Antoni Gaudi, Lebbeus Woods, and Peter Zumthor complete the top 10 and are often used in explicit architectural context given the strong red bar.

Figure 2(d) shows the links between keywords and the most likely connected term. We analysed this by predicting with the Word2Vec for each keyword on the left the most probably co-located word on the right, weighted by probability. Interesting combinations here are links between interior-design, floor-plan-drawing, architecture-visualization, cathedral-gothic, or swimming-pool. From this it is possible to build an of auto-complete function for architectural queries.

Figure 2(e) shows the mean length of queries depending on whether they got upscaled, remastered or left in draft mode. A draft mode image is of low image size, so users will normally upscale or remaster them if they like one of the variants. It is notable that for the medium upscale options as well as the remastered version the mean query length increases above 35 terms per query in comparison to 30 terms for draft mode queries. We also manually classified the most frequent 150 terms into the categories: style, content, quality. It is notable that for the upscale and refined queries, the percentage of style terms increases significantly from 6.6 % to 8.3 %.

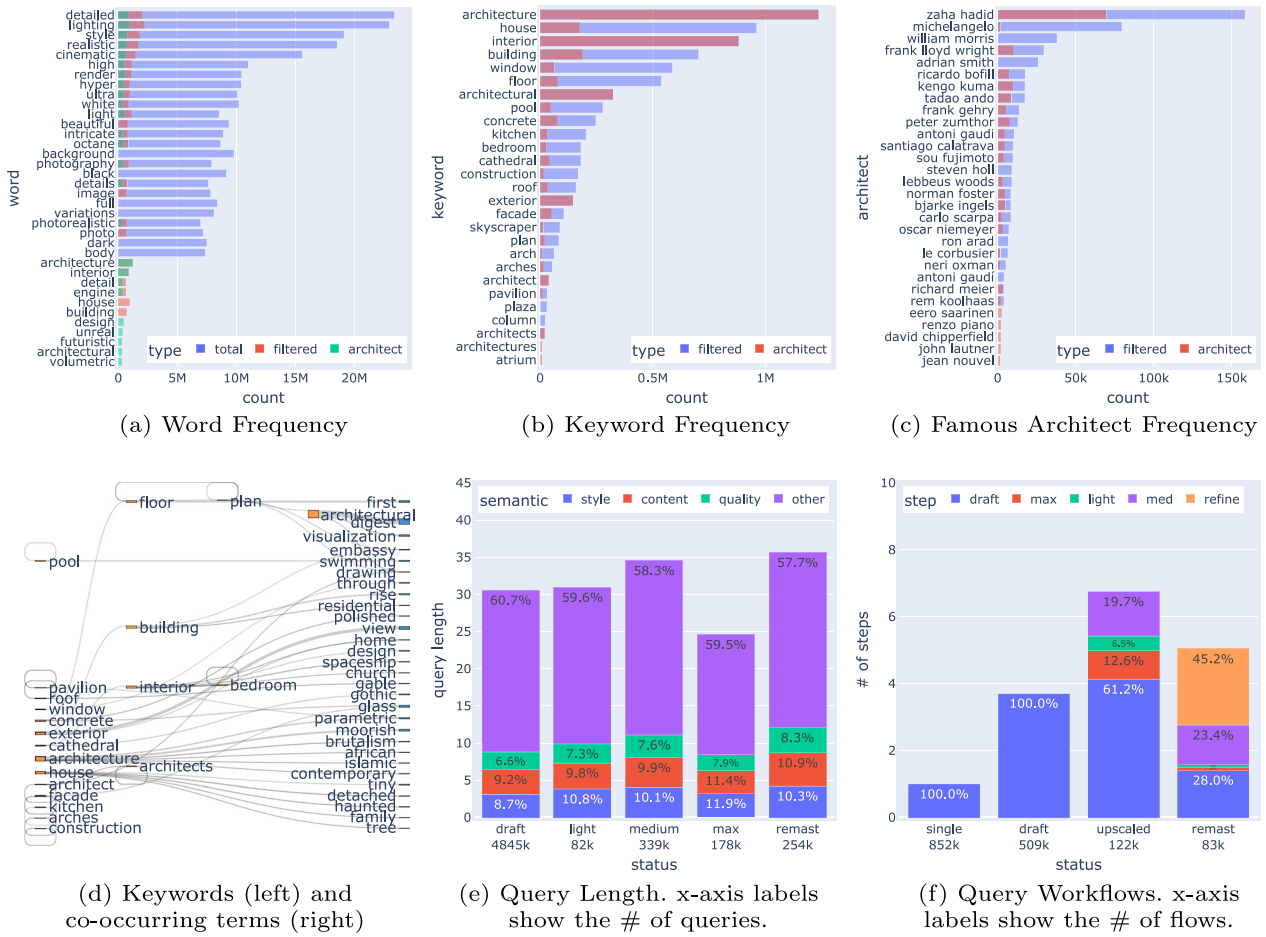
Figure 2(f) shows the mean number of iterations needed to develop a query. We classify a query as iteration if the same user is rerunning the same or extended

[5] Keywords are not listed in Fig. 2(b) due to low count: balcon, basilica, battlement, buttress, gable, hvac, latticework, livingroom, minaret, panelling, pavilion, plinth, rotunda, spire.

[6] https://www.heritage.nf.ca/articles/society/architectural-terms.php.

[7] https://en.wikipedia.org/wiki/Glossary_of_architecture.

[8] https://en.wikipedia.org/wiki/List_of_architects.



(a) Word Frequency (b) Keyword Frequency (c) Famous Architect Frequency

(d) Keywords (left) and co-occurring terms (right) (e) Query Length. x-axis labels show the # of queries. (f) Query Workflows. x-axis labels show the # of flows.

Fig. 2 Results from analysing Midjourney queries

query within 30 min. 54 % of all unique queries (single, 852k) are run only once. The other half of the queries are improved in multiple iterations. Queries that remain in draft mode require about 3.7 steps. 7.8 % of the queries are good enough to be upscaled require about 6.75 steps in total. They are upscaled after 4.1 draft steps into different variants (light, medium/beta, max). 5.3 % queries that are remastered take about 5.1 iterations. They have only 1.4 draft mode queries, but 1.2 remastering steps, and 2.3 final upscale steps.

This analysis illustrates that users do not come-up with perfect queries from scratch. We can derive multiple insights:

- 6.7% (5.7 million) of all queries are related to architecture;
- "architecture" is the most popular keyword, and "interior" is six times more popular than "exterior" as a keyword;
- only 13.1 % of all unique queries upscaled or remastered;
- these queries are usually refined in more than 4 iterations in mean;
- these queries usually contain more keywords specifying style and quality.

5 Case studies

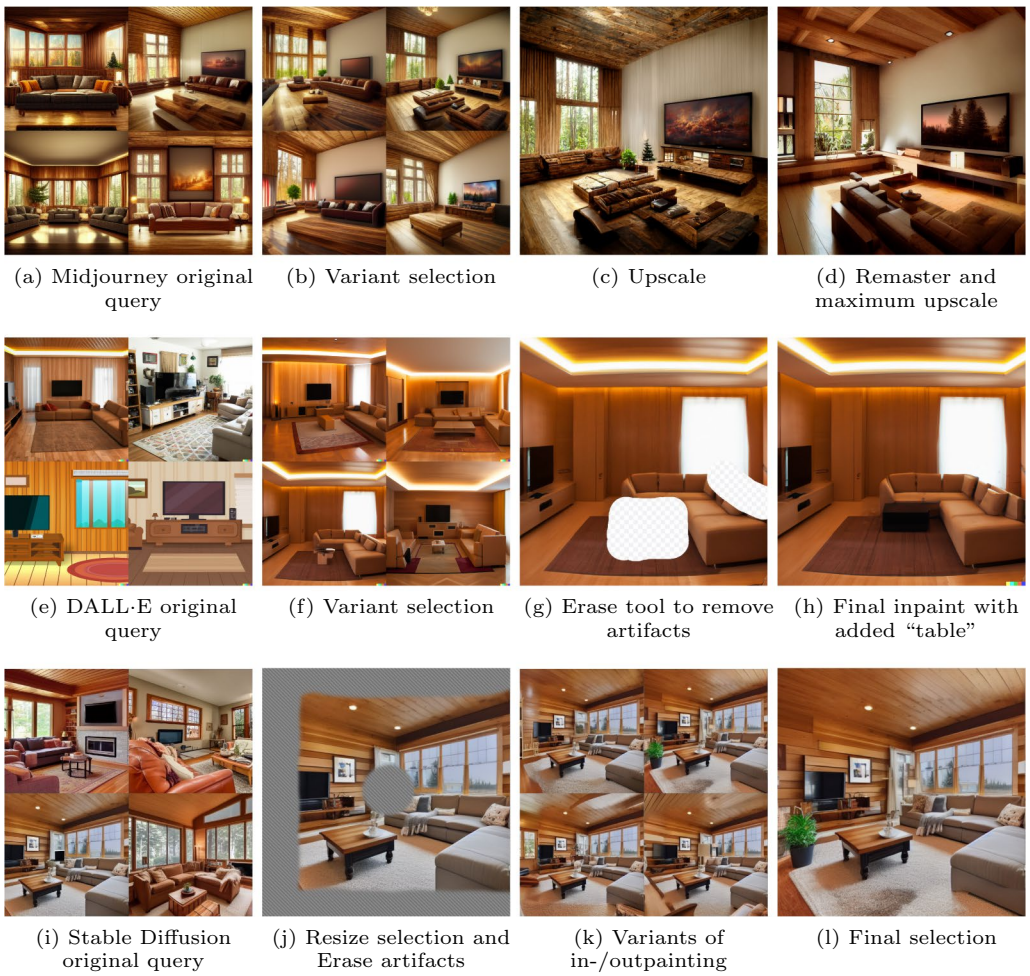
In this section, we present some refined workflows for architects to utilize the strengths of all three AI art platforms. These workflows are based on the learning of our analysis and by a large number of experiments to identify the workflows leading to the best results. It should be noted that Midjourney v3 is in use here, which raises the importance of the remastering step. Remastering can usually be skipped as of model version 4. As we have identified in the analysis, users rarely run only a single query and gain a perfect result—instead they iterate many times. This requires a good understanding of effective workflows to avoid dead ends and come to adequate results quickly.

5.1 Interior design—comparing the models

First, we will look at how the models perform purely on their own. The example will be an interior-design scenario. We start with a simple query without any special command of a platform. On all three we try to generate a high-quality rendering of a room using the prompt "cozy living room, wood paneling, television, large sofa, natural light, lived in, realistic, full view". We developed this prompt by watching common prompting patterns in the Midjourney query data and testing out different iterations across models until we arrived at that final version. Each comma is considered as topic separator by the AI art platforms. Thus, we are asking for an image of a (i) cozy living room that (ii) has wood paneling; (iii) contains a TV; (iv) a large sofa, etc. With this we ensure that the resulting image should contain similar elements across platforms.

Midjourney starts out with several results that are stylized or of strange perspective, visible in Fig. 3(a, b). The first upscale of the chosen interior design in (c) greatly improves material quality and overall detail, but the central sofa remains as an incoherent form in the centre of the image. Anytime the normal image output does not attain a sufficient level of cohesion and realism, we can invoke the "remaster" step, shown upscaled in (d). However, even this last result contains smaller perspective errors, which are difficult to fix without intervention through manual image editing.

The DALL·E 2 first results in Fig. 3(e) shows that it struggles to correctly response to the "realistic" term in the query. Once one of the two realistic variants is picked, the next variant generation step in (f) creates more useful results. To remove perspective or coherence errors we mask out certain areas of the image in (g) to generate inpaint variants. The final variant is shown in (h). Of note is that even the inpainted regions react correctly to the previously established lighting of the scene. This result is of good quality, but cannot be upscaled any further within the web interface.



(a) Midjourney original query (b) Variant selection (c) Upscale (d) Remaster and maximum upscale

(e) DALL·E original query (f) Variant selection (g) Erase tool to remove artifacts (h) Final inpaint with added "table"

(i) Stable Diffusion original query (j) Resize selection and Erase artifacts (k) Variants of in-/outpainting (l) Final selection

Fig. 3 Minimal workflow for Midjourney (**a**–**d**), DALL·E 2 (**e**–**h**), and Stable Diffusion (**i**–**l**) for the given query

Stable Diffusion starts out in Fig. 3(i) with much stronger results than its two contenders, generating images that incorporate the "realistic" and "lived in" aspect of the query very well. These rooms look like inhabited and not like artificial renderings. However, all images seem like close-ups of a proper interior view. Which is why we do not just use inpainting in (j) to fix errors, but also add additional canvas space for outpainting. This results in some quite incoherent variants for the outpainted areas. After some additional iterations the variant in (l) was selected as the best.

Overall, Stable Diffusion performs best in this scenario. All of its first variants were realistic and showed no real deficiencies in the later steps. Only during the final outpainting stage it was necessary to manually smooth the transitions. Midjourney generated a final result of similar quality, however offered a weaker beginning selection, with not all images containing all prompted elements (e. g. missing TV) and frequent perspective errors and other visual faults.

5.2 Exterior design—combining the models

The best results can be archived by combining the strengths of all three models and knowing their specific command keywords. For example, one of the first hints that DALL·E shows new users is the information that the keyword "digital art" can significantly improve many prompts, which is deeply related to the data it was trained on.

In the case of Midjourney, refinement starts as a process of including and removing certain phrases within the prompt to get as close to a desired style as we can. These phrases can be very convoluted. It often helps to include the kinds of modelling software that would create the desired kind of image (like "octane render" or "cinema3D") or even quality signifiers like "top 10 on artstation" into the prompt. A much more directed way to influence queries is the use of word weights, image references and parameters, which add additional parameterization to the prompt system.

In contrast, DALL·E and Stable Diffusion provide more control over changes with their in- and outpainting tools. Once we understand these tools and strengths of the platforms, we can combine them into a more flexible workflow. In the following ideation workflow, we will start with a Midjourney prompt to create a desired scene. While Stable Diffusion tends to generate better looking first results, Midjourney is a strong contender once the remaster step is done, and with its workflow excels at free-form ideation, which makes it the most appropriate starting point for an exterior design. From the Midjourney remaster stage, we will then refine any errors or undesired results through in-/outpainting in DALL·E and Stable Diffusion to attain the targeted result. An overview of the workflow is provided in Fig. 4 and we will explain it along the results in Fig. 5. The workflow highlights how the different image editing steps can be combined to get the best results independent of the AI tool used. The specific models most appropriate for each step may change with newer versions. The logic behind the workflow is targeting the best image quality, by: (1) generating the image; (2) upscaling it; (3) extending the canvas; (4) finally editing details with inpainting.

Figure 5(**a**–**c**) shows the beginning stages of an idea as generated in Midjourney. The query that led to this particular result was "single-family home with garden, full exterior view, modern architecture, photo, sunlight". Multiple keyword arrangements and weights on different terms were tried before this result was selected, remastered and then upscaled. It is also possible to start with one or more reference images, though that technique was omitted here.

The result was then uploaded to the DALL·E, where it was outpainted in Fig. 5(d) to create a wider viewing angle and subsequently inpainted in Fig. 5(e, f) to replace unwanted details like the cables hanging in the air or the differently colored windows.

After unsuccessful attempts to add a paved path from the sideway to the entrance through inpainting in DALL·E, we transferred image (f) to Stable Diffusion.

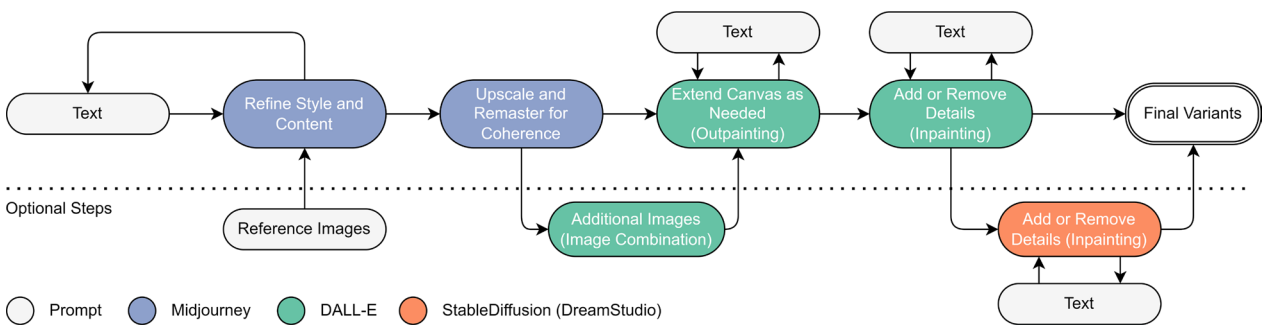


Fig. 4 The proposed combined workflow over Midjourney, DALL·E and (optionally) Stable Diffusion

We erased the walkway and replaced it with inpainting using a modified query without a garden reference (as the walkway would often be overgrown by plants) and adding "paved" early in the prompt (as earlier keywords are weighted higher). This resulted in the prompt "single-family home, paved between sidewalk and door, full exterior view, modern architecture, photo, sunlight" with the result in image (g).

The speed of the process allows to develop this design together with a client. He may also suggest major changes like adding a second story or some other roof element. This can easily be accomplished via inpainting by erasing the roof and part of the sky and a slightly changed prompt specifying the style of the new image. Figure 5(h, i) shows the result for the Stable Diffusion query "single-family home, two stories, clear blue sky, curved roof, full exterior view, modern architecture, photo, sunlight" after the roof and central area of the sky have been erased and the canvas has been extended upwards to give more room for roof elements. Note that some latent effects like the tree branches reaching into the image are hard to avoid.

5.3 Common limitations

Working with current AI models is an trial-and-error process. They rarely present perfect results on first try, but expect the user to pick the best variants and refine their prompts multiple times. This may not be straight forward, but giving a group of architecture students the task to draw a design with similarly rough specifications would also result in many variants and would be way more time consuming.

Nonetheless, many variants that users explore fail entirely. This is evidenced by the high number of single step queries without upscaling in our Midjourney analysis in Fig. 2(f). Some of the common failure cases an architect would encounter while using these models are shown in Fig. 6. Case (a) shows the result for a floor plan query. It does well in imitating the style of bold lines for walls and thin lines for objects, but is completely nonsensical on closer look. This is due to the fact that AI art tools replicate the style, but have no semantic understanding of the meaning of the lines in a floor plan. Case (b) shows the result of a query with multiple specific technical terms, which are also somewhat ambiguous in

(a) Midjourney original query (b) Variants (c) Remastered and Upscaled

(d) DALL·E outpainting of (c) (e) Erased parts (f) Inpainted

(g) Stable Diffusion inpainted (h) 2^{nd} story variant 1 (i) 2^{nd} story variant 2

Fig. 5 Refinement and variant generation in Midjourney (**a**–**c**), DALL·E 2 (**d**–**f**), and Stable Diffusion for a walkway (**g**) and a second story (**h**, **i**)

(a) "architectural floor plan of a modern single-family home, ground floor"

(b) "red brick building facade with a dutch clock gable and a bow window"

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

Fig. 6 Example failure cases

themselves. The specific architectural element "bow window" was turned into a window with a bow over it. The "clock gable" was represented by a different gable element with a clock below it. Case (c) shows a query for a landscape design with five buildings, which invokes the common problem that these AI tools are just bad at counting and complex spatial arrangements beyond foreground and background.

6 Conclusion

In this paper we looked at AI art generation tools in their applicability to architecture and civil engineering. We compared three of the currently available AI art platforms and identified use cases that can be tackled now or are soon to be unlocked. To understand how users are already using these tools we analyzed millions of queries providing some insights on how users iterate. Finally, we presented two workflows, for interior and exterior design, with the latter combining the strengths of the different platforms.

The various use cases shown in this paper illustrate the strong potential for AI tools in architecture. The AI platforms still struggle with more complex prompts, usually due to missing semantic understanding of the image content. A floor plan for example is not just a collection of lines. These lines carry semantic and contextual information, like that they form walls enclosing a room with a door to get in. As we already have Building Information Models (BIM) that provide this semantic information it is just a matter of time that new diffusion models will arrive that are trained specifically on these data sets, and it is likely that they will be able to fulfil all requirements from Table 1.

Nonetheless, the high number of 5.7 million queries with architectural context that we identified show that the tools are already adopted. In the coming months and years these platforms will further improve. We can already see workflows across tools that converge toward Fig. 4. In the end, single platforms will deliver the full workflows for use cases like ideation, collages, build and style variants that will drastically improve productivity and creativity. It is thus likely that these tools will first be adopted for brain storming sessions with clients and for competitions. We observed that the designs created by current AI tools tend toward organic forms, openwork facades and complex arrangements that break out of the common minimalistic modern design. With the ongoing research on automated evaluation of structure dynamics and in the field of additive and robotic construction technologies, more and more of these designs are becoming structurally and financially possible. *This may form a perfect storm situation leading to a new generation of architecture styles based on AI-generated designs.*

Author contributions
JP: Ideation and manuscript, Midjourney Study, Writing Sects. 1, 3, 4. MB: Use Case Experiments, Writing Sects. 2, 3, 5. Both authors reviewed and edited all sections as well as read and approved the final manuscript.

Funding
No funding was received for conducting this study.

Data availibility
The Midjourney datasets generated and analysed during the current study are not publicly available for data privacy reasons. Code for retrieving the data can be made available from the corresponding author on reasonable request.

Declarations

Competing interests
The authors have no financial or proprietary interests in any material discussed in this article.

Received: 8 February 2023 Revised: 28 June 2023 Accepted: 31 July 2023
Published online: 17 August 2023

References

Borji, A. (2022). Generated faces in the wild: Quantitative comparison of stable diffusion, midjourney and DALL-E 2. arXiv preprint http://arxiv.org/abs/2210.00586.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., & Agarwal, S. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877–901.

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Nips, 33*, 6840–6851.

Ho, J., Salimans, T., Gritsenko, A.A., Chan, W., Norouzi, M., & Fleet, D.J. (2022). Video diffusion models. ICLR workshop on deep generative models for highly structured data.

Kawar, B., Zada, S., Lang, O., Tov O, Chang, H., Dekel, T., Mosseri, I., & Irani, M. (2022). Imagic: Text-based real image editing with diffusion models. arXiv preprint http://arxiv.org/abs/2210.09276.

Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., & Van Gool, L. (2022). Repaint: Inpainting using denoising diffusion probabilistic models. CVPR (pp. 11461–11471).

Luo, S., & Hu, W. (2021). Diffusion probabilistic models for 3D point cloud generation. CVPR (pp. 2837–2845).

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. Nips (Vol. 26).

Nichol, A., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., McGrew, B., Sutskever, I., & Chen, M. (2022). Glide: Towards photorealistic image generation and editing with text-guided diffusion models. ICML.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G. (2021). Learning transferable visual models from natural language supervision. ICML (pp. 8748–8763).

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with CLIP latents. arXiv preprint http://arxiv.org/abs/2204.06125.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022, June). High-resolution image synthesis with latent diffusion models. CVPR (p. 10684–10695).

Saharia, C., Chan, W., Chang, H., Lee, C., Ho, J., Salimans, T., Fleet, D., & Norouzi, M. (2022). Palette: Image-to-image diffusion models. ACM SIGGRAPH (pp. 1–10).

Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J., & Norouzi, M. (2022). Image super-resolution via iterative refinement. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Seneviratne, S., Senanayake, D., Rasnayaka, S., Vidanaarachchi, R., & Thompson, J. (2022). DALLE-URBAN: Capturing the urban design expertise of large text to image transformers. International Conference on Digital Image Computing: Techniques and Applications.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., & Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. ICML (pp. 2256–2265).

Tevet, G., Raab, S., Gordon, B., Shafir, Y., Cohen-Or, D., & Bermano, A.H. (2022). Human motion diffusion model. arXiv preprint http://arxiv.org/abs/2209.14916.

Zeng, X., Vahdat, A., Williams, F., Gojcic, Z., Litany, O., Fidler, S., & Kreis, K. (2022). LION: Latent point diffusion models for 3D shape generation.

Zhou, L., Du, Y., & Wu, J. (2021). 3D shape generation and completion through point-voxel diffusion. CVPR (pp. 5826–5835).

Publisher's Note