Spatial interpretation in a text-to-image accelerated architectural design process.

VISSERS-SIMILON, E. and DOUNAS, T.

2024

Spatial Interpretation in a Text-to-image Accelerated Architectural Design Process

Investigating diffusion models in sketch, preliminary and final design stages

Elien Vissers-Similon[1], Theodoros Dounas[2]
[1,2]Universiteit Antwerpen
[1]{elien.vissers-similon|theo.dounas}@uantwerpen.be

The introduction of diffusion models – artificial intelligence models originally used as text-to-image generators – poses several questions for the architectural design process. Firstly, could diffusion models enhance those design processes? Secondly, what is the relationship between innovative image generators and traditional methods of representation derived from projectional geometry? This paper studies the results of such an accelerated design process by 76 masters students of architecture, which lasted 6 weeks and covered the sketch, preliminary and final design stages. We define spatial interpretation moments as the inflection points the during human-machine interaction when the designer translates 2D images into 3D spatial design concepts. The spatial interpretation moments mostly occur in the transition from sketch to preliminary design and during preliminary design. Spatial interpretation moments' inherent opportunity is to use diffusion models both as a communication and a design tool to rapidly test spatial design intentions. This paper showcases examples of the captured spatial interpretation moments regarding the designer's ability to actively design and the impact of spatial dimensions, spatial composition and spatial abstraction. Moreover, this paper suggests the use of annotations to capture spatial interpretation moments for future research and proposes boundaries to investigate the relationship between diffusion models and other methods of representation.

Keywords: *Diffusion Models, Midjourney, Enhanced Design Process, Human-Machine Interaction, Generative Design.*

INTRODUCTION

Diffusion models – artificial intelligence (AI) techniques initially used as text-to-image generators – have been adopted by architects rapidly. The 2024 RIBA AI survey reveals that 41% of architecture practices in the UK uses AI for at least the occasional project (Malleson, 2024) since OpenAI released its first diffusion model Dall-E (Ramesh et al., 2021) in 2021. Diffusion models allow architects to generate new images based on text, reference images, or a combination of both. Moreover, plugins such as ControlNet (Zhang et al., 2023) facilitate masking, sketch-to-image and the integration of control images such as depth maps. Recent classification of various AI techniques (Vissers-Similon et al., 2024a) shows that diffusion models are one of the most accessible AI techniques to architects, and hold great potential for applications in early architectural design stages.

This poses the question whether rapid and intuitive image generation could be a catalyst for a paradigm shift in architectural design processes, of which design acceleration and methods of representation are crucial aspects. This paper studies a designer-focused experience within a complete diffusion-powered design sprint: how does the designer interact with the computer during the sketch, preliminary and final design stages? We define 'interpretation moments' to investigate human-machine interaction in each design stage: inflection points during the design process when designers are forced to make design decisions based on the received feedback from the machine.

BACKGROUND

Studies into the use of diffusion models are often limited to the sketch design stage and show creative and ideation benefits of integrating rapid image generation in the architect's early workflow (Dortheimer et al., 2023, Guida, 2023, Çiçek et al., 2023). We conducted a six-week diffusion-powered design sprint to simulate a complete workflow: from design brief to sketch design, to preliminary design, to final design and deliverables. The results have been analysed in two parts.

The first part of the research data focused on the sketch design stage and identified textual and visual interpretation moments in a diffusion-powered process, which forced the designer to make early design decisions concerning the programme, materiality, context, geometry and atmosphere (Vissers-Similon and Dounas, 2024). However, design decisions concerning geometry remained superficial and design decisions about spatial composition or structure rarely occurred, indicating that more research into the designer's spatial translation – from 2D generated images to 3D designed concepts – is needed.

This paper investigates the second part of the research data which looks into the complete process (sketch, preliminary and final stages) to assess when this spatial translation occurs, along with its relationship to traditional methods of representation. We define spatial interpretation moments as the human-machine inflection points that facilitate this spatial translation from 2D generated images to 3D designed concepts. We aim to answer the following research questions:

1. What defines the spatial interpretation moment in a diffusion-powered design process and when does it occur?
2. What are the opportunities and challenges of the spatial interpretation moment?
3. Do diffusion models stimulate spatial interpretation moments when used in parallel with traditional representational methods?

RESEARCH METHODOLOGY

Scope
The conducted experiment is part of a larger research project that aims to determine *how* architects could use diffusion models to improve or optimize their design process workflows. In order to do so, we focus on the human-machine interaction through advanced simulation and data analysis methods from the field of Advanced Engineering Informatics (Hartmann and Trappey, 2020): we investigate the behaviour of human designers and virtual designers (diffusion models) in a complex system. Several experiments will be conducted, each experiment learning from previous experiments to define new set-up and boundary conditions. The first phase of experiments is conducted within architecture curricula, and uses students as semi-experienced designers. This allows us to carefully prepare for a workshop with practicing architects in the second phase. This greatly increases the speed at which experiments can be prepared (which is crucial in a rapidly evolving field) and reduces the risk of not being able to deduct valuable conclusions from a larger workshop with practicing architects.

Set-up & Assumptions

The conducted six-week diffusion-powered design sprint is the first experiment within the project and mainly serves an explorative purpose.

The research data is collected from 76 master students of architecture divided into groups of 4, at the architecture department of the University of Antwerp. The design teams were each assigned mission statements to develop a building of medium complexity. Examples include community building, reconsidering education, re-wilding, care close by, boomer house renovation and others. The complete assignment including mission statements is available together with the research data. The design teams were asked to compile their own design brief based on the mission statement, e.g. the mission statement 'care close by' led to a design brief for a mobile care station, shown in Figure 1. Then, the teams had six weeks to create a project file up to tender phase, i.e. the final design, bill of quantities, specifications, budget and a timeline for building the project.

Figure 1
'care close by'
mission statement
resulting project

Each team was asked to use Midjourney, a diffusion model that allowed the research team to capture all produced prompts and images in dedicated discord servers. The use of Midjourney was obligatory during the sketch design stage, recommended during the preliminary design stage and optional during the final design stage. Designers were free to use any other design tools or methods of representation. The collected data include the design briefs, Midjourney prompts and images, submitted project files and answers to a reflection questionnaire about the designers' experiences, used methods of representation and the overall process. The sketch design consists of translating the design brief into initial ideas, traditionally visualised with sketches or other methods of representation. The preliminary design is the stage when the main parameters are defined: total area per function, number of storeys, global geometry, general materiality, etc. The final design stage contains detailed design decisions which lead to the bill of quantities, specifications, budget and project timeline.

Methods

The semi-experienced designers received weekly guidance and feedback sessions from experienced practicing architects who teach at the University of Antwerp, packaged as 'client meetings'. We enforced certain parameters of the process, such as the mission statements, the desired output and the use of Midjourney as a diffusion model. The designers were discouraged in using other diffusion models, as the focus of the experiment lies on the workflow and not diffusion model performance. Our methods include qualitative and quantitative analyses of the collected data. The paper consists of three data processing stages.

The first stage investigates the questionnaire's answers concerning the overall experience of using Midjourney in an architectural design process, such as what is the perceived added value of using Midjourney in each design stage and would the designers use Midjourney in later projects. The qualitative analysis of these answers is juxtaposed to the quantitative usage of Midjourney in each design stage. This allows us to define the spatial interpretation moments, assess when they would

occur and identify some of their opportunities and challenges.

The second stage analyses the design output and answers to the reflection document to conclude which methods of representation were used, and why, to pinpoint the impact of diffusion models on traditional 19th-century methods of representation. Designers were asked whether or not they used Midjourney in each design stage, which other methods of representation they used and which of those were most important to them.

The third stage analyses the visual design output and interviews the experienced practicing architects who served as the 'client' in order to identify more opportunities of the spatial interpretation moments.

Experiment

Firstly, two thirds of the designers indicated that they would use Midjourney again, which indicates that diffusion models might have permanently found their way into the architectural design process. The technology is still upcoming and new but its easy interface and intuitive nature promise a steep increase in use over the next few years.

What defines the spatial interpretation moment and their challenges?

During the sketch design stage, designers were obligated to use Midjourney. During the preliminary design, the use of Midjourney was recommended and during the final design stage using Midjourney was optional. Afterwards, we asked the designers whether they experienced an added value to using Midjourney during the sketch design and whether or not – and why – they used Midjourney during the preliminary and final design stages. A summary of the responses is visualised in Figure 2. A majority of designers (79%) experienced an added value of using Midjourney during sketch design. This is in line with previous research that investigated the benefits of diffusion models in the earliest design stages (Dortheimer et al., 2023, Guida, 2023, Çiçek et al., 2023). This is also reflected in the fact that 2/3 of the designers stated they would use Midjourney in later

projects, out of which 40% mentioned it would specifically be for the sketch design stage. The added value was mostly found in the speed, inspiration, ideation and facilitating internal communication within the design team. Other value was found in external communication with the client and the different focus on the designer's role in selecting the right images. Designers that did not experience an added value cite that the results were "too spectacular", the software could not reproduce an exact image of what the designers had in mind, and that the diffusion model was unsuitable for before-and-after images of renovation briefs. On average, designers stated to have spent 23% of their time during the 6 week design experiment on the sketch design. A summary of the responses about the added value of Midjourney in the sketch design is visualised in Figure 3.
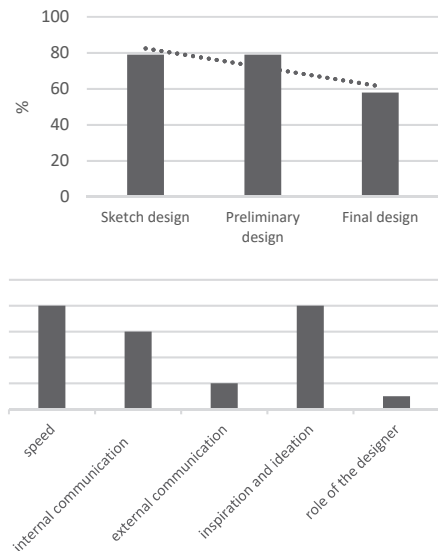
Figure 2
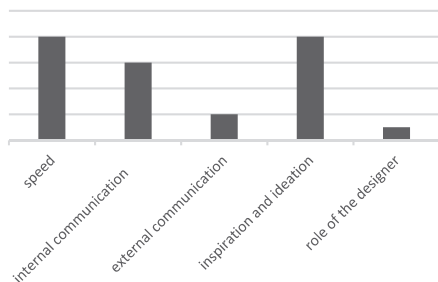Added value and usage of Midjourney during each design stage

Figure 3
Added value of Midjourney in sketch design

During the preliminary design, an equal but different majority of designers (79%) made use of Midjourney. Out of that group, 70% of experiences were reported to be positive, whilst 30% were reported to be

negative. In any case it is clear that diffusion models alone are not capable to complete a full preliminary design assignment. The positive experiences reiterated the added value of the sketch design – speed, inspiration and ideation, internal and external communication – whilst also appreciating the possibility to generate images showing the "atmosphere" and "design intention" rather than a 100% accurate and representative design drawing. This means that, for example, the generated images could show design intentions for the shape of the geometry, whilst the materiality and programme would not be represented accurately in that same image. The negative experiences expressed difficulty when designing more specific ideas or client demands, hallucinations that are structurally unbuildable, and they rejected the fact that the software did not generate an image that was 100% representative for the whole design, i.e. an accurate geometry, surroundings, materiality, programme, etc. all in one generated image. We note that the negative experiences were often related to designers who had first drawn floorplans and were

then trying to generate matching images. On average, designers stated to have spent 36% of their time during the 6 week design experiment on the preliminary design.

During the final design, slightly over half of the designers (58%) used Midjourney. Out of that group, most experiences were positive, referencing again to the speed of image generation and the efficient external communication with the client. Designers also mentioned focusing on interior images and adapting previously generated images, as most spatial design decisions had already been made. The negative experiences mainly referred to the already established difficulty of trying to generate a very specific image, or in other words, the fact that the generated image is not 100% accurate with all of the designer's wishes.

On average, designers stated to have spent 41% of their time on the final design stage of the 6-week experiment.

Table 1 summarizes the (dis-)advantages, the average time spent for each design stage, how many designers used Midjourney and how many had

Table 1
Usage of
Midjourney during
the complete
design experiment

|	SKETCH DESIGN	PRELIMINARY DESIGN	FINAL DESIGN
time during experiment	23%	36%	41%
Midjourney usage	100% (obligatory)	79% (recommended)	58% (optional)
added value, positive experience	79%	70%	90%
advantages	speed inspiration & ideation internal communication external communication important role of designer	speed inspiration & ideation internal communication external communication design intention	speed external communication focus on interior images adapt existing images
disadvantages	results 'too spectacular' not 100% accurate unsuitable before-and-after	not 100% accurate unbuildable hallucinations	not 100% accurate
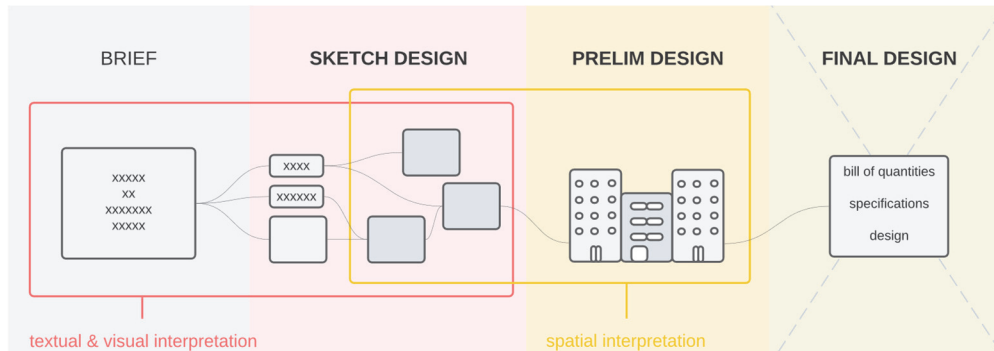
positive experiences. Diffusion models seem to have a greater potential for advantages during the sketch and preliminary design stage. The use during the final design stage is more contested as less designers opted to use the software and less advantages were identified. Taking into account that designers spent the least amount of time in the sketch design stage (23%), the preliminary design stage (36%) emerges as an interesting design stage for further research.

Since the first part of the research data (Vissers-Similon and Dounas, 2024) already indicated that design decisions concerning geometry and spatial composition or structure remained superficial during the sketch design, and the fact that the data above suggests that very few spatial design decisions are made during the final design (due to a focus on external communication and interior images), we can deduct that spatial interpretation moments – moments when the designer interacts with the computer and is required to make spatial design decisions: translating 2D images into 3D design concepts – mostly occur at the start of, or during the preliminary design stage.

We therefore redefine the spatial interpretation moments as moments, in the transition from sketch to, or during, preliminary design, when the designer interacts with the computer to translate 2D images into 3D spatial design concepts. The position of spatial interpretation moments within the full workflow is visualised in Figure 4, alongside

previously established interpretation moments (Vissers-Similon and Dounas, 2024). The associated challenges are (i) to produce 100% accurate consistent results with other representational methods and (ii) to translate hallucinations to buildable structures.

Do diffusion models stimulate spatial interpretation moments when used in parallel with traditional representational methods?
As stated before, we recommended that the designers use Midjourney during the preliminary design stage, and about 79% of them did. Additionally, the designers were free to use any other methods of representation during the preliminary design stage. A majority of designers chose floorplans to represent their design work, positioning floorplans at the same usage as Midjourney. Other, less frequently used methods included sections, sketches, diagrams, textual descriptions, reference images and 3D models.

However, when asked which methods of representation were considered the most important for the design team, floorplans were clearly deemed the most important and they were referred to as the "base documents to extract the deliverables from". There was a clear distinction with the usage of the other methods, where the second place was reserved for 3D models, and the third place was awarded to Midjourney. The distribution of used and

most important representational methods – as perceived by the designers – is summarised in Figure 5.

■ used methods ■ most important methods
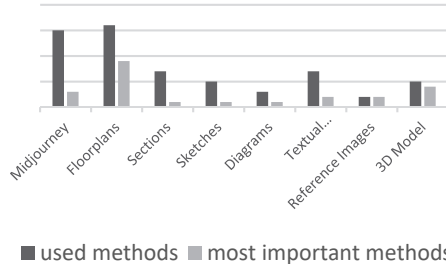
Midjourney was praised as a tool to rapidly communicate design intentions during the preliminary design stage, but did not seem to create any impact on the traditional 19th century orthographic methods of representation, i.e. designers still made spatial decisions by using floorplans, sections and 3D models. Midjourney was used as a tool to communicate about spatial intentions rather than a tool to make spatial decisions.

Do spatial interpretation moments entail more opportunities?

As the explicit data was unable to capture any spatial design decisions from the designer, we delved into the implicit research data by re-examining the visual output whilst interviewing the practicing architects who served as the 'clients' and provided the designers with feedback during the experiment. The visual output and interviews enabled us to reconstruct four spatial interpretation moments.

The first identified spatial interpretation moment pertains to *the designers' ability to actively design* context-bound architecture (materiality, geometry) as opposed to the generic and often 'spectacular' architectural images that are often produced by diffusion models. The design team started out with the generated images represented in Figure 6, but as their project was supposed to be

located in Valencia, they made use of photographs as reference images (shown in Figure 7) to redesign the exterior of their building, creating a design more suitable and realistic for the Spanish context, and continuing the assignment towards a more buildable project file and deliverables. Through carefully selecting the reference images, the design team took active control over the generated spatial output. The final result is visualised in Figure 8.

The second identified spatial interpretation moment pertains to *the impact of spatial dimensions*. The design team generated two images of the same community centre's main space, one image when the space is empty and one image when the space is furnished, shown in Figure 9. The rapid image generation has helped the design team to correctly estimate the required dimensions of the main space, when drawing the floorplans afterwards. A critical note is that this example also showcase the challenge of structurally translating unbuildable hallucinations, as the generated image shows a structural grid of beams, but no visible load bearing columns or walls. The design team resolved this by

introducing structural columns alongside the curtain wall.

The third identified spatial interpretation moment pertains to *the impact of spatial composition* and is visible in one of the rewilding project images shown in Figure 10. The image of the small interior garden helped the design team to improve the spatial composition of multiple wings of their building, in order to create small and qualitative outdoor spaces. Similar to the previous example, it also guided the design team when establishing the adequate spatial dimensions – depth, width and height – of the exterior space.

The fourth identified spatial interpretation moment pertains to *the impact of spatial abstraction* and is also visible in the rewilding project images shown in Figure 10. Here, the abstract images

assisted the designers to adapt their design-point-of-view: they were confronted with the fact that rendered images are not always 'final' and do not always need to be 100% representative.

DISCUSSION

Spatial interpretation moments are defined as moments in the transition from sketch to preliminary design or during preliminary design, when the designer interacts with the computer to translate 2D images into 3D spatial design concepts. The challenges are (i) to produce 100% accurate consistent results with other methods of representation and (ii) to translate hallucinations to buildable structures. The first challenge (i) can be flipped into an opportunity: if designers adapt their mindset and do not expect the generated images to be 100% accurate design representations, a diffusion-powered workflow allows rapid intuitive generation and testing of various design intentions. The second challenge (ii) is also not problematic, as the experiment showed that as long as the designer is critical of the generated images, they will be able to spot unbuildable hallucinations and offer design alternatives – or be inspired by the hallucinations and attempt to incorporate them into a realistic design proposal, e.g. by using computational design software to rationalise a seemingly complex design. During the preliminary design stage – when most spatial decisions were made – designers were free to use any representational methods. This decision has led to the successful completion of the design task at hand, but also allowed the designers to fall back on what they knew, i.e. making spatial decisions based on familiar, traditional, 19th century methods of representation. Hence, no spatial interpretation moments were captured straight from the designers and the collected data does not say much about the relationship between traditional methods of representation and diffusion models as a design tool.

We recommend that future research into spatial interpretation moments attempts to capture them straight from the designer by asking the designers to annotate the generated images with their design

Figure 9
Results community building project

Figure 10
Results of various rewilding projects

intentions: this ensures that the designer reflects on the image and what design parts or aspects of the image he or she wants to explore further. More research on the occurrence, nature and criteria of the spatial interpretation moments should be conducted in order to assess whether or not diffusion-powered architectural design workflows are indeed faster and/or more efficient during the preliminary design stage. We recommend that future experiments focus only on the preliminary design and set strict guidelines on which representational methods are and are not allowed. The next experiment in our research project (Vissers-Similon et al., 2024b) makes use of such strict guidelines in order to learn more about the relationship between traditional representational methods and diffusion models, when both are used in parallel.

As for the final design stage, slightly more than half of the designers elected to use Midjourney during the final design of the experiment and most experiences were reported to be positive, which indicates more in-depth research into the usage of diffusion models in the final design might be beneficiary as well.

CONCLUSIONS

The designers' experiences confirmed that diffusion models do have a place in the architectural design workflow. This paper offers a framework to conduct future research by introducing spatial interpretation moments. We define these moments as the inflection points during human-machine interaction when the designer translates 2D images into 3D spatial design concepts. The spatial interpretation moments occur in the transition from sketch to preliminary design and during preliminary design. More research into their occurrence is recommended in order to establish a feasible workflow for architects to integrate diffusion models into their practice. The main question is how the full potential of diffusion models could be used in an architectural design process, whilst still using the full potential of traditional methods of representation as

well, since it is clear that diffusion models alone are not capable of conducting a complete design assignment from brief to deliverables.

This paper identifies challenges and opportunities when integrating diffusion models into the architectural design workflow. The challenges are (i) to produce 100% accurate consistent results with other methods of representation and (ii) to translate hallucinations to buildable structures. The opportunities are to use diffusion models as a communication tool, as well as a design tool to rapidly test (spatial) design intentions. Examples of the captured spatial interpretation moments pertain to the designer's ability to actively design and the impact of spatial dimensions, spatial composition and spatial abstraction. Moreover, this paper suggests the use of annotations to capture spatial interpretation moments for future research and proposes boundaries to investigate the relationship between diffusion models and other methods of representation.

OPEN ACCESS / RESEARCH DATA

The research data, which includes figures, the assignment and the questionnaire, can be found in the following online repository:

https://github.com/ElienVissers/diffusion-models-accelerated-architectural-design-process

REFERENCES

Çiçek, S., Turhan Haskara, G. & Özkar, M. 2023. Reconsidering Design Pedagogy through Diffusion Models.

Dortheimer, J., Schubert, G., Dalach, A., Brenner, L. & Martelaro, N. 2023. Think AI-side the Box! Exploring the Usability of Text-to-Image Generators for Architecture Students.

Guida, G. 2023. Multimodal Architecture: Applications of Language in a Machine Learning Aided Design Process.

Hartmann, T. & Trappey, A. 2020. Advanced Engineering Informatics - Philosophical and methodological foundations with examples from civil and construction engineering. Developments in the Built Environment, 4.

Malleson, A. 2024. Artificial Intelligence Report: RIBA AI Survey Findings. Royal Institute of British Architects.

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M. & Sutskever, I. 2021. Zero-Shot Text-to-Image Generation. International Conference on Machine Learning, Vol 139, 139.

Vissers-Similon, E., De Walsche, J. & Dounas, T. 2024a. Classification of artificial intelligence techniques for early architectural design stages. International Journal of Architectural Computing.

Vissers-Similon, E. & Dounas, T. 2024. Textual and Visual Interpretation in a Text-to-image Accelerated Architectural Design Process. Architecture Across Boundaries. Suzhou, China.

Vissers-Similon, E., Verbruggen, S., De Walsche, J. & Dounas, T. 2024b. Spatial Interpretation in a Diffusion-powered Architectural Design Workflow. Conference for Artistic and Architectural Research Valencia. Valencia.

Zhang, L., Rao, A. & Agrawala, M. 2023. Adding Conditional Control to Text-to-Image Diffusion Models.