ESTOA **28** / pp. 121-135 / DOI: 10.18537/est.v014.n028.a09

Research Article
2025 July - December

Semiotics-based prompt engineering for architectural text-to-image generation processes Ingeniería de prompts basada en semiótica para procesos de generación de imágenes a partir de texto en arquitectura

09

ŞULE TAŞLI PEKTAŞ
OSTIM Technical University, Turkey
sule.taslipektas@ostimteknik.edu.tr

BILGE SAĞLAM
OSTIM Technical University, Turkey
bilge.saglam@ostimteknik.edu.tr

ABSTRACT Text-to-image generative AI tools have gained significant attention in the architectural community; however, they are currently being used by trial-and-error with simple textual inputs. This is largely due to the lack of established frameworks for crafting prompts that yield semantically rich architectural outputs. This paper proposes using semiotics as an analytical method facilitating text-to-image generation processes. Two experiments were conducted to investigate the effects of semiotic analysis and adding context modifiers to prompts on the relevancy of outputs of three mainstream text-to-image generation tools (DALL-E, Midjourney, and Stable Diffusion). The results indicate the effectiveness of the proposed method and reveal opportunities and limitations of current text-to-image generative models in architecture. It is concluded that a human-centered approach to Human-AI interaction is needed to overcome issues regarding control, transparency, and data quality.

RESUMEN Las herramientas generativas de IA de texto a imagen ya han atraído la atención de la comunidad arquitectónica; sin embargo, actualmente se utilizan mediante prueba y error con entradas textuales simples. Esto se debe en gran medida a la falta de marcos establecidos para la creación de indicaciones que generen resultados arquitectónicos semánticamente ricos. Este artículo propone el uso de la semiótica como un método analítico que facilita los procesos de generación de texto a imagen. Se llevaron a cabo dos experimentos para investigar los efectos del análisis semiótico y la adición de modificadores contextuales a las indicaciones en la relevancia de los resultados de tres herramientas principales de generación de texto a imagen (DALL-E, Midjourney y Stable Diffusion). Los resultados indican la efectividad del método propuesto y revelan tanto oportunidades como limitaciones de los modelos generativos de texto a imagen actuales. Se concluye que se necesita un enfoque centrado en el ser humano para la interacción Humano-IA, a fin de superar problemas relacionados con el control, la transparencia y la calidad de los datos.

Recibido: 8/10/2024
Revisado: 03/01/2025
Aceptado: 28/01/2025
Publicado: 29/07/2025

KEYWORDS architectural design, generative ai, text-to-image generative models, prompt engineering, semiotics

PALABRAS CLAVE diseño arquitectónico, ia generativa, modelos generativos texto-imagen, ingeniería de prompts, semiótica

Cómo citar este artículo/How to cite this article: Pektas, S. T. & Saglam, B. (2025). Semiotics-based Prompt Engineering for Architectural Text-to-Image Generation Processes. *Estoa. Revista de la Facultad de Arquitectura y Urbanismo de la Universidad de Cuenca, 14*(28), 121-135. https://doi.org/10.18537/est.v014.n028.a09

e - ISSN: 1390 - 9274
ISSN: 1390 - 7263

ESTOA 28 / Vol 14 / 2025
e - ISSN: 1390 - 9274
ISSN: 1390 - 7263

1. Introduction

Recent years have witnessed considerable growth in generative Artificial Intelligence (AI) applications, particularly in the Machine Learning (ML) sub-field (Huang et al., 2021). Several ML applications utilizing natural language processing and text-to-image generation technology have been developed lately. An AI text-to-image generator uses a machine learning (ML) technique called artificial neural networks to receive textual input, process words, and generate an image. The earlier generations of such models relied on Generative Adversarial Networks (GANs) which involve two neural networks competing with each other. One network, the generator, is responsible for creating images, while the second network, the discriminator, is used to determine whether or not the images are real or fake. The rivalry between the two networks improves the efficiency of the systems rapidly and enables satisfactory results (Goodfellow et al., 2014). Recently, diffusion models have outperformed GANs and become state-of-the-art image generators (Dhariwal and Nichol, 2021). Diffusion models are ML systems that are designed to remove noise from images. They are trained on millions of text/image pairs; thus, as a response to a text prompt i.e. a short descriptive text, the system generates a new image that matches the prompt (Radford et al., 2021).

Diffusion-based text-to-image generation tools have matured at an unprecedented rate in recent years. Although their first generation emerged in the early 2010s, they attracted massive public attention after 2021 when the new tools became capable of generating complex and realistic outputs (Steinfeld, 2023). As in many other fields, these tools have increasingly been used in the field of architecture due to their potential to enhance the creative aspects of design processes. They are being adopted by professional architects, since they enable quick transformation of textual descriptions into detailed visual representations (Autodesk, 2024). Paananen et al. (2023) explored how generative AI tools support creativity in the early stages of the architectural design process and discussed that participants found them useful for generating new design ideas and developing concepts more efficiently than traditional methods. The study's findings highlighted the importance of integrating such tools into the architectural workflow to foster innovation and facilitate the ideation process.

Despite the growing interest in generative AI, the nature of human-AI co-creation has rarely been studied. As Liu and Chilton (2022) put forward, human-AI interaction during text-to-image generation has two aspects. On the one hand, users are able to input anything and have access to numerous generations. On the other hand, they also "must engage in brute-force trial and error" in order to reach desirable outputs. Due to the iterativeness and the experimental nature of the process, the practice and skill of writing prompts is called "prompt engineering". Prompt engineering is an emerging research area in Human-Computer Interaction (HCI) and there are currently few academic studies on it. Moreover, the majority of existing studies have focused on natural language processors, while studies on text-to-image generation have been rare. Additionally, such studies do not tackle the problem from a disciplinary perspective.

Our study aims to fill these research gaps and investigate text-to-image generation processes in architecture. The motivation of the study is the vision that in the near future, the generative capabilities of ML systems will reach a level of creating holistic designs. Therefore, in our work, prompt engineering was approached not from the popular "something in the style of artist/movement" manner, but semantically rich textual input, which is more suitable for creating architectural narratives, were studied. This paper proposes semiotics (the systematic study of signs and symbols and their interpretation) as an analytical method facilitating text-to-image generation processes. It is widely acknowledged that meaning in architecture is created with words as well as forms (Psarra, 2009). Therefore, understanding and improving the performance of text-to-image generative AI for architectural narratives is important.

Within this perspective, the research questions of the study are presented below:

1. What are the opportunities and challenges of text-to-image generation processes in architecture?
2. What are the effects of semiotic processing of prompts on the relevancy of outputs?
3. What are the effects of adding a context modifier (building type and style) to the prompts on the relevancy of outputs?

In order to answer these questions, two experiments were designed and implemented. The excerpts from the novel *The Fountainhead* (Rand, 1943/2015) were utilized in the experiments because the book includes samples of architectural narrative (Section 2.2. presents the arguments for selecting the specific parts of the novel for the analyses).

The definitions of two buildings from the book were extracted and classified as either denotative (literal) or connotative (associative) statements. While denotations were fed into the text-to-image generation tools as they were, connotations were utilized both in unprocessed and semiotically-analyzed form. In order to understand the effect of adding a context modifier to the prompts, in each of the experiments the procedures were repeated once more after adding the context to each prompt in each of the experiments. The results were tabulated and compared. The evaluations of the visual outputs and the assessment of generative processes provided valuable insights into the use of generative AI in architecture. To the best of our knowledge, this study is the first to propose semiotics as an analytical tool aimed at facilitating prompt engineering and human-AI co-creation.

The paper is organized as follows. The following sub-sections of the introduction review the recent developments in generative AI in architecture, text-to-image generation technology, and prompt engineering.

The second section explains the methodology of the study. The third section presents the results. The fourth section includes a discussion of the findings and the final section addresses the limitations of the study and makes suggestions for further research.

1.1. Generative AI in architecture

Using image data as input has been widely researched since Goodfellow et al. (2014) introduced the Generative Adversarial Networks (GANs). The method was practiced in the architectural field before the text-to-image generators' popularity in the literature and professional fields (Horvath and Pouliou, 2024). Currently, GANs utilize a wide range of built environment-related data, including GPS, street view images, 3D models, architectural drawings, and building performance models, and are applied at buildings, district, and city levels (Wu et al., 2022). At the building level, GANs were used for many purposes, such as generating spatially optimized designs (Nauata et al., 2020; Luo & Huang, 2022; Sun et al., 2022), restoring old building facades (Zhao et al., 2020), providing inspiration for architects (Chen & Stouffs, 2021) and collaborating with students by integrating datasets with their sketching skills (Akcay Kavakoglu et al, 2022). Chaillou (2019), introduced the ArchiGAN (a custom version of GAN) and used it to generate housing floor-plan layouts in three steps of footprint massing, program repartition, and furniture layout (Chaillou, 2020).

The Convolutional Neural Networks (CNNs) is another well-known deep learning architecture that draws inspiration from live species' innate visual perception mechanisms. Such systems are used for image recognition, processing, as well as classification and proved to be especially useful for extracting features from input data (Gu et al., 2018). Although the development of early CNNs dates back to the late 1980s (LeCun et al., 1989), they gained widespread attention from architectural community after 2020. For example, del Campo et al. (2021) used the Convolutional Neural Networks (CNNs) to create new architectural styles through hallucinations. They presented methodologies for a posthuman design ecology to expand the creative process by transferring architectural features from 2D to 3D forms and "dreaming" urban textures in the landscape design for the project "Robot Garden" (del Campo et al., 2021).

Apart from the experiments for 3D formation and building layout generation, the research-through-design methodology is one of the areas that can nourish the process of architectural creative thinking and design (Horvath and Pouliou, 2024). In a collaborative process of text-to-text, text-to-image, and image-to-image generation, Horvath and Pouliou (2024) followed three steps of curation, automation, and re-curation to examine the large dataset of architectural texts and workflows of the hybrid text and hybrid images in the design process of an architectural competition. Even though their study introduced an insightful and novel methodology, the systematical examination of the selected text was not provided in terms of semantic reading and evaluation. Our study focuses on this research gap and scrutinizes text-to-image generative AI. A short account of the developments in this technology is presented in the following section.

1.2. A brief review on text-to-image generation and prompt engineering

Early text-to-image generation models date back to the first half of the 2010s when developments in deep neural networks advanced (Baghadlian, 2023). However, the most important advances in the field were made after 2021. Radford et al. (2021) developed CLIP (Contrastive Language-Image Pre-training), a method for learning multimodal representations. OpenAI's DALL-E utilized CLIP technology and was released in January 2021. A newer version capable of generating more complex and realistic images, DALL-E 2, was announced in April 2022 and other commonly known generators Stable Diffusion and Midjourney was launched in Summer 2022. Although different techniques have co-existed during this period, the basic approach that has been adopted by popular text-to-image generation systems including DALL-E, Stable Diffusion, Midjourney and others is CLIP-guided diffusion. As the name implies, there are two components in this technique: CLIP and diffusion. The former is a system that generates images to match a given prompt, while the latter generates images through a "de-noising" process. CLIP-guided diffusion was invented by Crowson in 2021 based on earlier work at Stanford University and UC Berkeley (Sohl-Dickstein et al. 2015; Ho et al., 2020). The researchers at these institutions developed

ESTOA 28 / Vol 14 / 2025
e - ISSN: 1390 - 9274
ISSN: 1390 - 7263

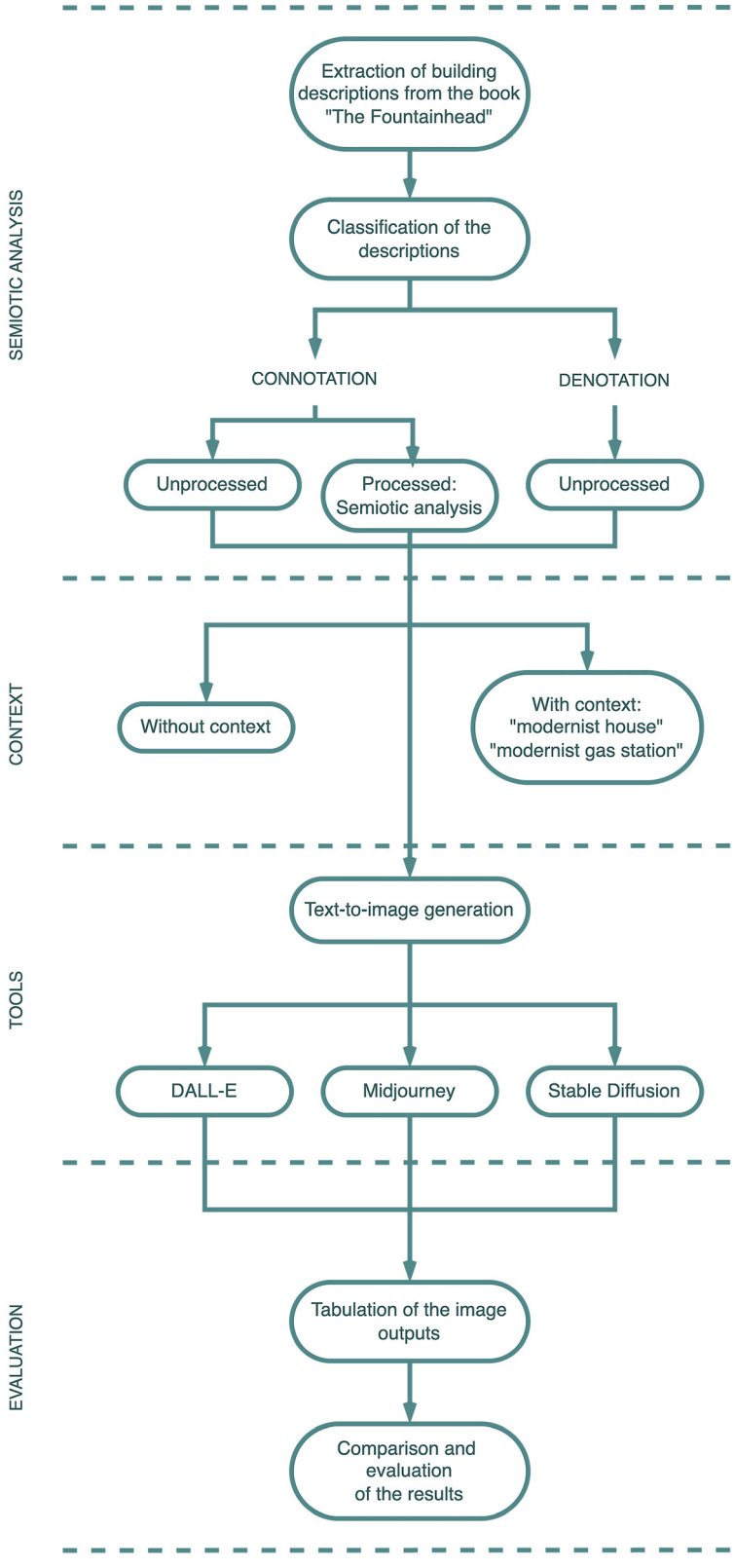


Figure 1: Research methodology diagram. (2025)

a neural network-based system to restore structure in noisy data after systematically destroying structure. In this way, the noise in the data was omitted. Following the development of text-to-image models, text-to-video platforms such as Runway (Runway ML, 2024), Imagen Video (Ho et al., 2022), and Phenaki (Phenaki, 2023) have emerged to generate videos from text/image prompts.

Even this brief account of research milestones highlights the unprecedented speed of development. When this manuscript was being prepared, most of text-to-image generation models had already reached a level of photorealistic reality and a substantial number of enthusiasts and professionals had already experimented with these tools. As Steinfeld (2023) discussed, the developments in the area after Summer of 2022 influenced "architectural visual culture suddenly, severely, and seemingly out of nowhere." Researchers and practitioners now have to tackle the problem of writing effective textual inputs for optimum results, a process known as prompt engineering. The term was originally coined to define the practice of devising effective prompts for the natural language processor GPT-3 (Oppenlaender et al., 2023). It was later used for text-to-image generation processes as well, but the majority of the work in this area has remained on natural language models and came from non-professional users. Besides discussions on online forums and social media, some online templates have been developed to guide prompt-generation processes for image creation. These templates and guidelines include PromptSource (Bach et al., 2022), DALL-E Prompt Book (DALL-E Prompt Book, 2022), and the Traveler's Guide to the Latent Space (Smith, 2022).

Despite these efforts, scholarly work on the subject is still rare and just emerging. Liu et al. (2023) studied prompting methods in natural language processing and called the emerging paradigm "pretrain, predict, and prompt." Liu and Chilton (2022) conducted five experiments to investigate what prompt parameters can help users produce better outputs from text-to-image generation models and presented the results as design guidelines. Oppenlaender (2023) proposed a taxonomy of prompt modifiers i.e. keywords and key phrases added to the prompts to obtain better results. He defined six distinct categories of prompt modifiers for text-to-image generation: subject term, style modifier, image prompt, quality booster, repeating term, and magic term. Pavlichenko and Ustalov (2023) discussed that human input is necessary to determine the ideal prompt formulation and keyword combination in text-to-image generation and developed a human-in-the-loop method for discovering the most effective prompts. Hao et al. (2024) developed an automated prompt engineering system for text-to-image generation based on supervised and reinforcement machine learning. The authors reported that their system outperformed manual prompt engineering in terms of both automatic metrics and human preference ratings.

A review of previous studies on generative AI reveals that the research focus has been on simple prompts and none of the existing works studied semantically rich textual inputs like the ones commonly found in architectural narratives. Moreover, earlier work in the field relied either on categorizing prompts and offering more effective combinations; or utilizing human or machine reasoning to create better prompts.

Our study takes an innovative approach and employs a well-established structured technique to guide text-to-image generation whether performed by humans or machines. Furthermore, our study examines the effectiveness of the proposed method through empirical analysis, which has also been lacking in the majority of previous work. The section below explains the research methodology of the study in detail.

2. Methodology

This study employed an experimental research method. The experimental method focuses on the effect of a change, referred to as a "treatment" or "intervention." This approach entails designs using standardized procedures to hold all conditions constant except the independent (experimental) variable (Ross and Morrison, 2013). Our study aims to compare the outputs of text-to-image generators for denotation and connotation statements. Another aim is to understand the effects of semiotic analysis and adding a context modifier as an intervention. Therefore, a strict experiment design framework was implemented in this study wherein everything except the interventions was controlled. A diagram depicting the research methodology developed and used in the study is presented in Figure 1.

2.1. Tools

The study utilized three novel open-access diffusion text-to-image generators (DALL-E 3, Midjourney V6, and Stable Diffusion 3.0). DALL-E 3 is a developed version of DALL-E created by OpenAI. Midjourney is an AI tool produced by an independent research lab that uses Discord's cloud service with bot commands. Since Midjourney launched in July 2022, version 6 was released, and the analysis was based on this last version. Stable Diffusion 3.0 is the latest version of the Stable Diffusion model developed by Stability AI. These three text-to-image generation models are the most widely used ones and our study employed the latest versions of them.

2.2. Semiotic analysis and image creation processes

This study employs semiotics to investigate the formation of meaning in architectural narratives and to guide the text-to-image generation processes accordingly. Ferdinand de Saussure's (1916/2011) linguistic model forms the basis of the definition and applications of semiotics in many areas, including literature, anthropology, art, and architecture. According

to Saussure, semiotics deals with how meaning is created through the relationship between the signifier and the signified. A signifier is any expression that signifies such as text, image, gestures, etc. Signified is the concept that the signifier conveys. The signifier and the signified together constitute the sign, the smallest unit of meaning. About 50 years after Saussure's death, French semiologist Roland Barthes further elaborated Saussure's semiotic theory and recognized the difference between denotation and connotation (Barthes, 1964/1967). According to him, denotation is a sign's most basic or literary meaning, while the connotation is a sign's secondary, cultural meaning (Figure 2).

The novel *The Fountainhead* by Ayn Rand (1943/2015) was chosen as the source of material for the experiments because it includes semiotically rich examples of architectural narrative interwoven with denotations and connotations. The book proposes an alternative New York City within the perspective of objectivity and the modernistic view of architecture. The book's protagonist, Howard Roark, resembles the well-known modern architect Frank Lloyd Wright since Wright's buildings and the imaginary ones mentioned in the book possess similar characteristics (Berliner, 2007).

Following the procedure recommended by Cullum-Swan and Manning (1994), the authors conducted a semiotic analysis. Two interpretants carried over the process. Both interpretants were architect-academicians who were highly knowledgeable in modern architecture and had experience in semiotic analysis. The analysis included textual samples for two prominent buildings in the book: the Enright House and the Gowan Service Center. The interpretants thoroughly read the book, and semiotic signs referring to these particular buildings were extracted. Signifiers (denotation and connotation types) and their signified meanings are specified and tabulated (Figure 3 and Figure 4). Then, three different text-to-image software applications (DALL-E 3, Midjourney V.6, and Stable Diffusion 3.0) were used to generate building images according to the textual samples. All the textual material was fed into these tools. While denotation signifiers were entered into the systems as they are, connotation signifiers were entered both in their original syntax and in semiotically analyzed form in order to understand the effect of semiotic intervention (Figure 1). The outputs of this process are presented in the following section.

3. Results

3.1. Evaluation of the visual outputs

The visual outputs of the text-to-image generation processes were recorded as diagrams (Figures 5 to 8 show the results).

In the first step of the Enright House experiment, no context modifier was added. In this step, DALL-E and Midjourney produced similar results for denotation and un-processed connotation statements (Figure 5). The results included brutal imagery and some surreal components and deviated from what was expected. However, semiotic intervention improved the results to a great extent for both of the tools. On the other hand, Stable Diffusion created relevant images for all types of statements i.e. denotation, un-processed connotation and semiotically-processed connotation, while semiotic intervention increased the precision and detailing of the results. All of Stable Diffusion outputs and the images produced by DALL-E and Midjourney after semiotic processing referenced Modern architectural style which is the main architectural discourse of the novel "The Fountainhead" (Rand, 1943/2015).

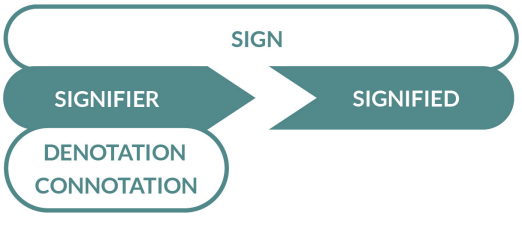


Figure 2: Barthes' definition of the basic elements of a sign. Diagram developed by the authors (Pektaş and Sağlam, 2025) according to Barthes 1964/1967

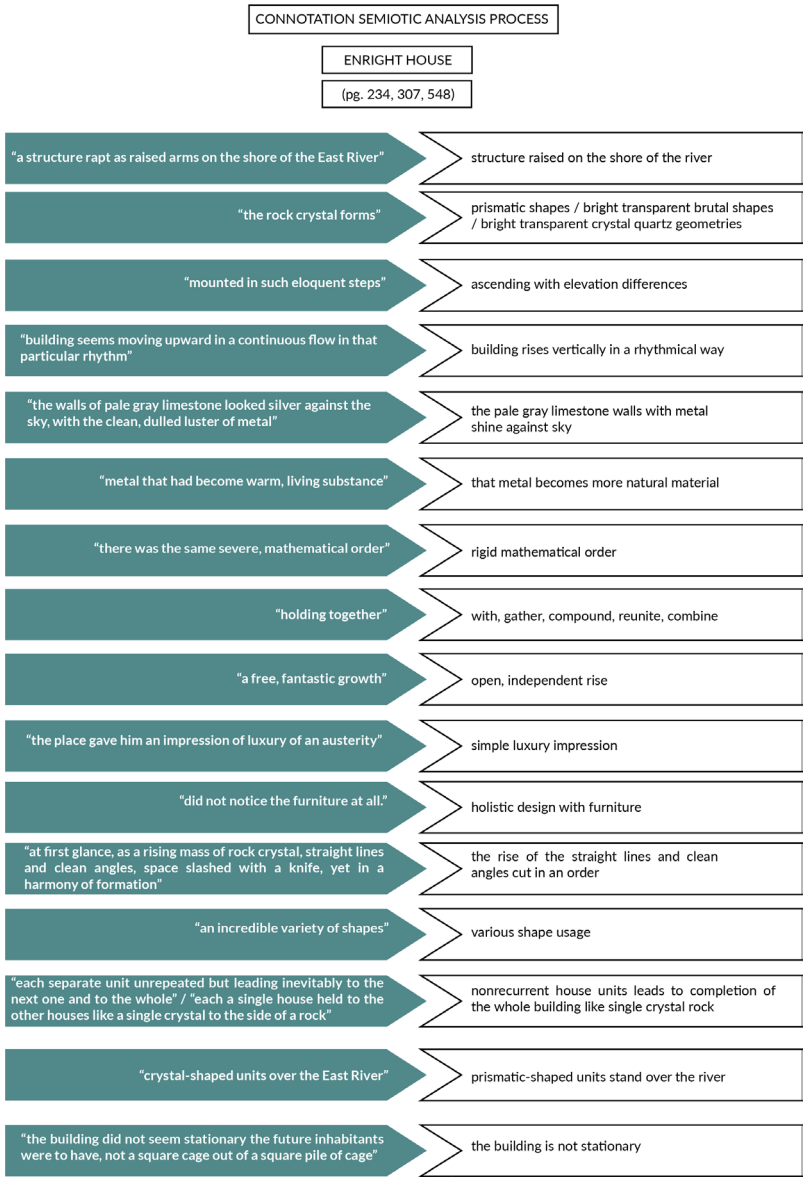


Figure 3: Semiotic analysis of connotation statements related to the Enright House. (2025)

This finding is notable because it shows that semiotic analysis increases the relevance of visual outputs to the extent that text-to-image generative AI can easily interpret the style of the associated building even when the style is not specified in the prompts.

In the second step of the Enright House experiment, the context of the building was added to the prompts as "modernist house." It was observed that this intervention improved the outputs of all tools to varying degrees (Figure 6). Midjourney was the most sensitive tool to the context modifier and enhanced its results greatly. The surreal components in the images were removed, and the results completely fitted within the framework of "modernist house" definition. Stable Diffusion produced context-relevant results as it did with the no-context prompts at the previous step, but its results were more clearly defined in this case with detailed depictions of architectural elements. The tool which was less sensitive to the style modifier was DALL-E. It produced similar results for denotation and semiotically-processed connotation statements and some rough results for un-processed connotation statements.

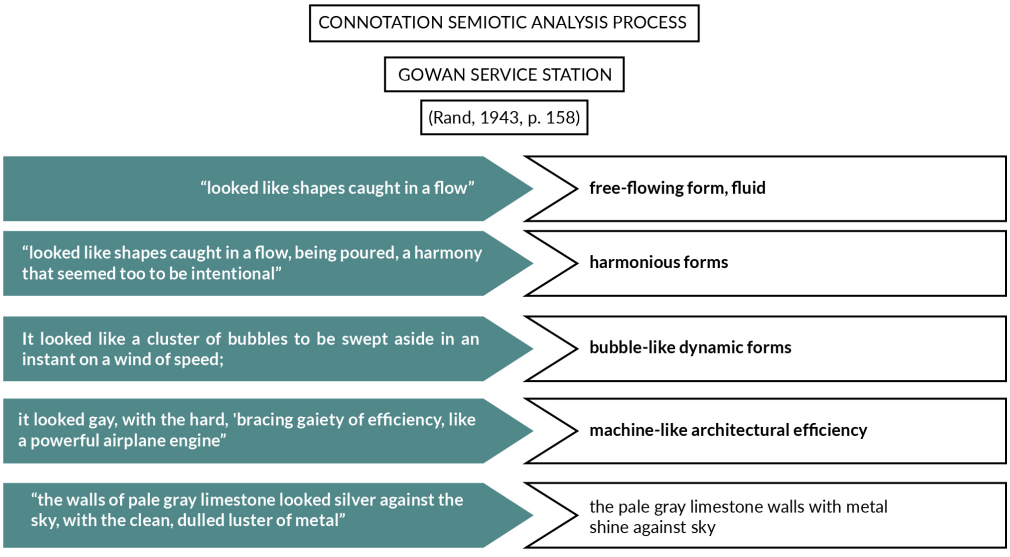


Figure 4: Semiotic analysis of connotation statements related to the Gowan Service Center. (2025)

In the second experiment, textual samples related to "Gowan Service Center" were employed. When the prompts were entered without context in the first step, the denotation statements led to visual results which seem to be consistent with the modernist gas station portrayals in the book for all generative models (Figure 7). However, it should be noted that in this category, Midjourney added some surrealist/fantasy elements and backgrounds to the depictions. The unprocessed connotation statements fed into Stable Diffusion and DALL-E resulted in abstract images that can be used as inspiration for conceptual design. On the other hand, Midjourney generated less abstract and literal results at this stage. This is probably due to Midjourney's literal interpretation of the phrase "like a powerful airplane engine." When semiotic analysis was applied to the connotation statements, it was observed that Midjourney and Stable Diffusion responded better to this intervention. The outputs produced by DALL-E for semiotically-processed and -unprocessed connotation statements were not significantly different.

When the context was added to the prompts as "modernist gas station" in the second step, the effects were similar to those that were observed in the first experiment (Figure 8). Again, Midjourney was the most sensitive to the style modifier. DALL-E produced the same results for denotation statements no matter if the context was specified or not. For un-processed connotation statements, it combined depictions of a modernist gas station with the unprocessed connotation outputs in the previous step in an awkward way, but semiotic intervention improved the results for connotation statements. Stable Diffusion produced context-relevant results for all types of interventions, but again semiotic intervention had a positive effect on the relevancy and creativity of the results. Although there was a dramatic improvement in the Midjourney's outputs for the connotation statements after the semiotic intervention when the context was not specified (Figure 7), the addition of the context had the most dominant effect on the connotation output images and almost erased the effect of semiotic intervention in the second step (Figure 8). While Stable Diffusion produced results that conform to the "modernist gas station" definition for all types of statements, Midjourney's results yielded more toward free-form futuristic architecture (Figure 8).

3.2. Evaluation of the text-to-image generation processes

Besides comparing the outputs of generative models under different conditions, our experiments provided insights into text-to-image generation processes. The main observed advantage of text-to-image generative AI was the rapidness of the tools. In architecture, producing visualizations is a cumbersome and time-consuming task. The AI-generated images are much faster to produce; one image can take just a few seconds to generate. The results have photo-realistic quality and much cheaper to produce compared to conventional means. However, there is also a dark side of this technology: the processes leading to meaningful results are not smooth and straightforward. Since the technology is still immature, there are many issues which need further work. The major process-related challenges of these tools observed in the experiments are lack of control, lack of transparency, and hallucinations.

First of all, it should be noted that current generative AI tools rely on the specification of intent as prompts through a chat interface. They receive a single prompt per trial which is assumed to describe the contents of the desired output image. When the prompt is implicit or conceptually loaded (as was the case for our un-

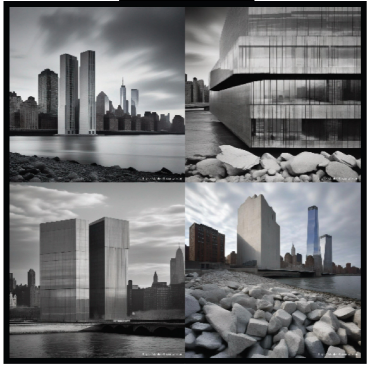
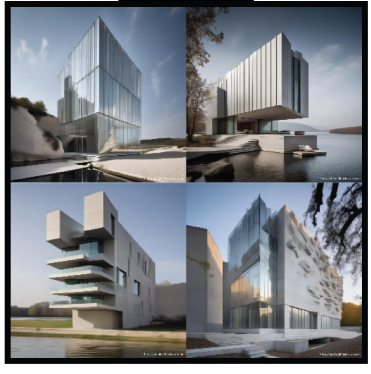
ENRIGHT HOUSE	ENRIGHT HOUSE	ENRIGHT HOUSE
Prompt: “the building stood on the shore of the East River, it was a structure on a broad space by the East River. The walls of pale gray limestone looked silver against the sky, with the clean, dulled luster of metal. At first glance, as a rising mass of rock crystal, straight lines and clean angles, space slashed with a knife, yet in a harmony of formation... an incredible variety of shapes, each separate unit unrepeated but leading inevitably to the next one and to the whole; each a single house held to the other houses like a single crystal to the side of a rock... crystal-shaped units over the East River... office, or home in the Enright House; a workroom, a library, a bedroom... only a clean sweep of space"	Prompt: “a structure rapt as raised arms on the shore of the East River, the rock crystal forms mounted in such eloquent steps that the building seems moving upward in a continuous flow in that particular rhythm, the walls of pale gray limestone looked silver against the sky, with the clean, dulled luster of metal that had become warm, living substance; at first glance, as a rising mass of rock crystal, straight lines and clean angles, space slashed with a knife, yet in a harmony of formation; there was the same severe, mathematical order, holding together a free, fantastic growth... the place gave him an impression of luxury of an austerity, did not notice the furniture at all.”	Prompt: “structure raised on the shore of the river, prismatic shapes (bright transparent brutal shapes / bright transparent crystal quartz geometries) ascending with elevation differences, building rises vertically in a rhythmical way, the pale gray limestone walls with metal shine against sky that metal becomes more natural material, rigid mathematical order combines with an open, independent rise, simple luxury impression, holistic design with furniture, the rise of the straight lines and clean angles cut in an order, various shape usage, nonrecurrent house units leads to completion of the whole building like single crystal rock, prismatic-shaped units stand over the river, the building is not stationary”
Denotation - Unprocessed without Context	Connotation - Unprocessed without Context	Connotation - Processed without Context
DALL-E	DALL-E	DALL-E
		
Midjourney	Midjourney	Midjourney
		
Stable Diffusion	Stable Diffusion	Stable Diffusion
		

Figure 5: Text-to-image generations of Enright House without context. (2025)

ENRIGHT HOUSE	ENRIGHT HOUSE	ENRIGHT HOUSE
Context: “modernist house” Prompt: “The building stood on the shore of the East River, it was a structure on a broad space by the East River. The walls of pale gray limestone looked silver against the sky, with the clean, dulled luster of metal. At first glance, as a rising mass of rock crystal, straight lines and clean angles, space slashed with a knife, yet in a harmony of formation... an incredible variety of shapes, each separate unit unrepeated but leading inevitably to the next one and to the whole; each a single house held to the other houses like a single crystal to the side of a rock... crystal-shaped units over the East River... office, or home in the Enright House; a workroom, a library, a bedroom... only a clean sweep of space"	Context: “modernist house” “a structure rapt as raised arms on the shore of the East River, the rock crystal forms mounted in such eloquent steps that the building seems moving upward in a continuous flow in that particular rhythm, the walls of pale gray limestone looked silver against the sky, with the clean, dulled luster of metal that had become warm, living substance; at first glance, as a rising mass of rock crystal, straight lines and clean angles, space slashed with a knife, yet in a harmony of formation; there was the same severe, mathematical order, holding together a free, fantastic growth... the place gave him an impression of luxury of an austerity, did not notice the furniture at all.”	Context: “modernist house” Prompt: “structure raised on the shore of the river, prismatic shapes (bright transparent brutal shapes / bright transparent crystal quartz geometries) ascending with elevation differences, building rises vertically in a rhythmical way, the pale gray limestone walls with metal shine against sky that metal becomes more natural material, rigid mathematical order combines with an open, independent rise, simple luxury impression, holistic design with furniture, the rise of the straight lines and clean angles cut in an order, various shape usage, nonrecurrent house units leads to completion of the whole building like single crystal rock, prismatic-shaped units stand over the river, the building is not stationary”
Denotation - Unprocessed with Context	Connotation - Unprocessed with Context	Connotation - Processed with Context
DALL-E	DALL-E	DALL-E
		
Midjourney	Midjourney	Midjourney
		
Stable Diffusion	Stable Diffusion	Stable Diffusion
		

Figure 6: Text-to-image generations of Enright House with context. (2025)

processed connotation statements), the outputs deviate largely from what was expected. On the positive side, this provides the user with some unexpected results which can enhance creativity and divergent thinking, but it can also be overwhelming and be regarded as a usability barrier.

Another issue related to text-to-image generation processes is the lack of transparency of the tools. The working mechanism of the generative AI is totally obscure to the user, so the user cannot interpret how the system reaches particular results. The systems are like black box; the user can only see the input and output, and what happens in between is incomprehensible. The controlled experimental setup of our study enabled the detection of behavior patterns of the systems; however, the lay user does not have the means for controlled inquiry, and reaching meaningful results with these tools for complex tasks can be really difficult.

Finally, our experiments showed that text-to-image generative models are prone to produce unrealistic, inaccurate, or fantasy outputs that are called "hallucinations" (Sahoo et al., 2024). AI models are trained on data, and they learn to make predictions by finding patterns in the data. If the training dataset is not accurate or contains biases, the model may learn to associate certain elements or features with unrelated concepts, causing it to "hallucinate." The common generative AI models are trained by large amounts of data already available on the Internet and many convenient datasets are dirty or biased (Whang et al., 2023). Crawford (2021) in her influential book "Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence" addresses the problem of "dirty data." Despite the widespread belief that algorithms are impartial and devoid of human bias, biased humans are the ones who generate the data sets and the code that determines how to utilize them (Crawford, 2021). Sahoo et al. (2024) discuss that the tendency of generative AI models to produce hallucinatory content represents the biggest obstacle to their widespread adoption in real-world scenarios, especially in areas where reliability and accuracy are crucial. Professional architectural applications entail realistic outputs and fit well to this definition. Therefore, minimizing hallucinations by improving data quality is a priority for architectural use of these tools. Generative AI systems specific to the architectural profession are yet to come and when such systems are developed, it will be important to use data that is relevant to the discipline.

4. Discussion

This study provided valuable insights into the use of generative AI for architectural purposes. Our experiments explored the impact of semiotic analysis and adding context to prompts on the relevancy of outputs of three mainstream text-to-image generation tools. The results demonstrated that while each tool has distinct internal mechanisms and behaves differently under various conditions, the application of semiotic analysis and contextual prompts generally enhanced the relevance of the outputs.

Distinct behavior patterns of the tools were observed during the experiments. For example, DALL-E responded positively to semiotical and contextual interventions, but its results overall were less relevant to the prompts, especially for un-processed connotation statements. Midjourney, on the other hand, was highly sensitive to the addition of contextual modifiers, producing significantly more detailed and articulate outputs when context was provided. Despite this, it tended more towards surreal and fantasy elements. Surprisingly, Stable Diffusion produced relatively more relevant results under all conditions, even in the absence of the modernist context in the prompts. The addition of the context resulted in more realistic architectural details with Stable Diffusion, as was the case with Midjourney. In all cases, the use of semiotic analysis improved the relevance of the outputs, particularly in situations where contextual information was missing.

This study showed that semiotics provides a structured framework for prompt engineering. By understanding and leveraging semiotic principles, designers and engineers can create more precise and effective prompts that guide AI models to produce images that closely align with intended concepts and meanings. The potential benefits of semiotics in prompt engineering can be discussed under three key areas:

Firstly, semiotics can increase the clarity and specificity of prompts. By analyzing the denotative and connotative meanings of words and phrases, users can select terms that precisely convey the desired visual elements. This reduces uncertainty and increases the likelihood that the image created will accurately represent the intended design.

Secondly, semiotics can help understand disciplinary, cultural and contextual nuances, which are critical to producing meaningful images. Sophisticated disciplines, like architecture, have their own discourses, terms, and jargon which can seem veiled to outsiders. By combining this understanding with rapid engineering, AI models can produce images that are not only stylistically accurate but also accurately resonate with the cultural context of the target audience. This is particularly important in global applications of text-to-image conversion, where cultural sensitivity (Eco 1976/1979) and accuracy are crucial.

Additionally, semiotics can help improve prompts iteratively. By analyzing the generated images and comparing them to the intended output, users can detect inconsistencies and refine prompts accordingly. This process involves semiotic analysis of both the text (prompt) and the resulting images, ensuring the alignment of semiotic elements in both. This iterative process may help fine-tune the AI model's understanding and rendering abilities, leading to increasingly better results.

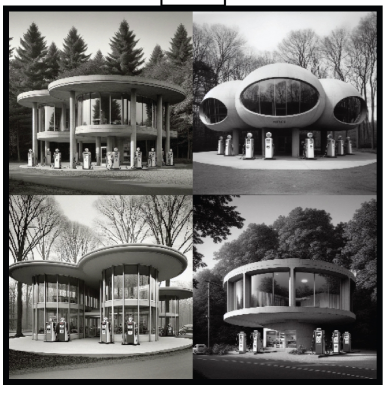
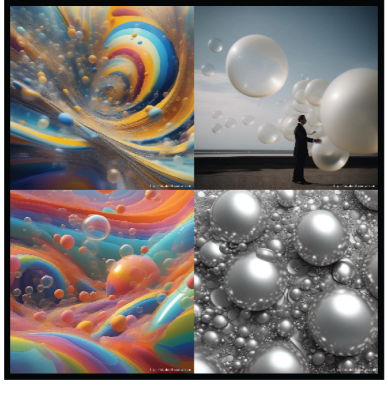
GOWAN SERVICE STATION	GOWAN SERVICE STATION	GOWAN SERVICE STATION
Prompt: "It stood on the edge of the Boston Post Road, two small structures of glass and concrete forming a semicircle among the trees, the cylinder of the office and the long, low oval of the diner, with the gasoline pumps as the colonnade of a forecourt between them. It was a study in circles; there were no angles and no straight line. It looked like a cluster of bubbles hanging low over the ground, not quite touching it."	Prompt: "looked like shapes caught in a flow, being poured, a harmony that seemed too to be intentional... It looked like a cluster of bubbles to be swept aside in an instant on a wind of speed; it looked gay, with the hard, 'bracing gaiety of efficiency, like a powerful airplane engine (Rand, 1943, p. 158)."	Prompt: "free-flowing form, fluid, harmonious forms, bubble-like dynamic forms, machine-like architectural efficiency"
Denotation - Unprocessed without Context	Connotation - Unprocessed without Context	Connotation - Processed without Context
DALL-E	DALL-E	DALL-E
		
Midjourney	Midjourney	Midjourney
		
Stable Diffusion	Stable Diffusion	Stable Diffusion
		

Figure 7: Text-to-image generations of Gowan Service Center without context. (2025)

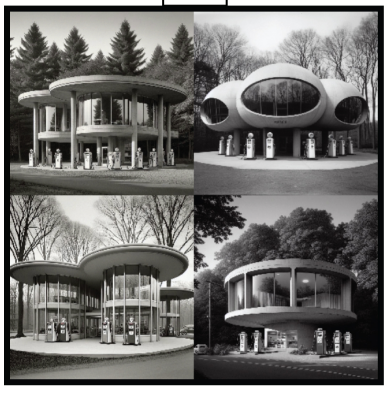
GOWAN SERVICE STATION	GOWAN SERVICE STATION	GOWAN SERVICE STATION
Context: “modernist gas station” “It stood on the edge of the Boston Post Road, two small structures of glass and concrete forming a semicircle among the trees, the cylinder of the office and the long, low oval of the diner, with the gasoline pumps as the colonnade of a forecourt between them. It was a study in circles; there were no angles and no straight line. It looked like a cluster of bubbles hanging low over the ground, not quite touching it.”	Context: “modernist gas station” “looked like shapes caught in a flow, being poured, a harmony that seemed too to be intentional… It looked like a cluster of bubbles to be swept aside in an instant on a wind of speed; it looked gay, with the hard, ‘bracing gaiety of efficiency, like a powerful airplane engine (Rand, 1943, p. 158).”	Context: “modernist gas station” Prompt: “free-flowing form, fluid, harmonious forms, bubble-like dynamic forms, machine-like architectural efficiency”
Denotation - Unprocessed with Context	Connotation - Unprocessed with Context	Connotation - Processed with Context
DALL-E	DALL-E	DALL-E
		
Midjourney	Midjourney	Midjourney
		
Stable Diffusion	Stable Diffusion	Stable Diffusion
		

Figure 8: Text-to-image generations of Gowan Service Center with context. (2025)

5. Limitations and future research directions

While this study provides important insights, it also has limitations that should be considered in further research. Our study employed an experimental research method in order to understand the relevancy of outputs from the three most eminent text-to-image generators when different categories of prompts were entered. Since these tools are still in early development, the results are not definitive but provide a foundation for future refinements.

This study highlights several problematic areas in the text-to-image generation process. The primary issues include the lack of control and transparency in the AI processes and the propensity for generating unrealistic or inaccurate outputs, known as "hallucinations." These challenges stem from the immature state of the technology and the reliance on large, often biased datasets. Addressing these issues is crucial for the broader adoption of AI in professional architectural applications. This paper suggests that semiotic analysis can enhance the generation process and emphasizes the need for high-quality, discipline-specific training data to minimize hallucinations and improve the reliability of generative AI systems in architecture.

Our paper introduces a novel qualitative analysis methodology using semiotic techniques to investigate a seminal literary text containing architectural descriptions. A broader range of architectural texts and historical or speculative documents can be subjected as case studies for future iterations to further evaluate the proposed framework's validity and versatility. Moreover, for future studies, the qualitative framework of this study can be enhanced with the involvement of the quantitative approaches, including semantic similarity and user preference scores. Using larger datasets and diverse scenarios, such as experiments with different architectural styles may also be suggested. In the end, using a larger dataset and mixed methodological process can enhance the results and make them more comprehensive in future studies.

Architectural design is an information-rich domain involving different kinds of activities. Therefore, the utility of generative AI tools should be investigated in relation to several tasks within the architectural design workflow in further studies. Moreover, our findings suggest a pressing need for the development of discipline-specific generative AI tools that are trained on high-quality, domain-specific datasets. This suggestion could open up a totally new track for future studies focused on enhancing the accuracy and relevance of AI-generated content for architecture.

Finally, in this study, semiotic analysis was conducted manually by human agents. However, this process could potentially be automated using machine learning techniques. Developing automated semiotic analysis tools can reduce biases in human interpretation and enhance data scalability. Therefore, further studies can expand on how semiotic analysis was automatized or integrated with prompt engineering more efficiently, further improving AI's ability to generate contextually appropriate and semantically rich images. The authors hope this study facilitates further exploration in this area.

Conflict of Interest. The authors declare no conflict of interest.

6. Bibliographic references

Akcay Kavakoglu, A., Almac, B., Eser, B. & Alacam, S. (2022). AI driven creativity in early design education – A pedagogical approach in the age of Industry 5.0, *Proceedings of the 40th International Conference on Education and Research in Computer Aided Architectural Design in Europe* (1), 133-142. https://doi.org/10.52842/conf.ecaade.2022.1.133

Autodesk. (2024). *How generative AI for architecture is transforming design*. https://www.autodesk.com/design-make/articles/generative-ai-for-architecture

Bach, S.; Sanh, V.; Yong, Z.; Webson, A.; Raffel, C.; Nayak, N.; Sharma, A.; Kim, T.; Bari, M.; Fevry, T.; Alyafeai, Z.; Dey, M.; Santilli, A.; Sun, Z.; Ben-David, S.; Xu, C.; Chhablani, G.; Wang, H.; Fries, J.; Maged, S.; Al-Shaibani; Sharma, S.; Thakker, U.; Almubarak, K.; Tang, X.; Radev, D.; Tian-Jian Jiang, M.; & Rush, A. (2022). Promptsource: An integrated development environment and repository for natural language prompts. *arXiv preprint arXiv.* https://doi.org/10.48550/arXiv.2202.01279

Baghadlian, S. (2023). *The complete timeline of text-to-image evolution.* Artificial Intelligence in Plain English. https://ai.plainenglish.io/the-complete-timeline-of-text-to-image-evolution-b63298234ed6

Barthes, R. (1967). *Elements of Semiology*. Hill and Wang.

Berliner, M. S. (2007). Howard Roark and Frank Lloyd Wright. In R. Mayhew (Ed.), *Essays on Ayn Rand's The Fountainhead* (pp. 41-64). Lexington Books, Plymouth, UK.

Chaillou, S. (2019). *AI + Architecture, Towards a New Approach* (Master's thesis). Harvard Graduate School of Design.

Chaillou, S. (2020). ArchiGAN: Artificial Intelligence x Architecture. In P. F. Yuan, M. Xie, N. Leach, J. Yao, & X. Wang (Eds.), *Architectural Intelligence* (pp. 117–127). Springer Nature Singapore.

Chen J. & Stouffs Z. (2021). From exploration to interpretation - adopting deep representation learning models to latent space interpretation of architectural design alternatives. In A. Globa, J. Van Ameijde, a. Fingrut, N. Kim, T.T.S. Lo (Eds.), *PROJECTIONS - Proceedings of the 26th CAADRIA Conference* - Volume 1, the Chinese University of Hong Kong and Online, Hong Kong, (pp. 131-140).

Crowson, K. (2021) *CLIP Guided Diffusion: Generates images from text prompts with CLIP guided diffusion.* https://colab.research.google.com/drive/1QBsaDAZv8np29FPbvjffbE1eytoJcsgA

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence.* Yale University Press.

Cullum-Swan, B. E. T. S., & Manning, P. (1994). Narrative, content, and semiotic analysis. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 463-477). Sage Publications.

DALL-E Prompt Book. (2022). *Guide to effective prompting.* https://dallery.gallery/the-dalle-2-prompt-book/

Del Campo, M., Carlson, A., & Manninger, S. (2021). Towards hallucinating machines-designing with computational vision. International *Journal of Architectural Computing*, *19*(1), 88-103. https://doi.org/10.1177/1478077120963366

Dhariwal, P., & Nichol, A. (2021). Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, *34*, 8780-8794. https://doi.org/10.48550/arXiv.2105.05233

Eco, U. (1979). *A theory of semiotics*. Indiana University Press.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, *27*, 2672-2680. https://doi.org/10.48550/arXiv.1406.2661

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X, Wang, G., Cai, J. & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, 354-377. https://doi.org/10.48550/arXiv.1512.07108

Hao, Y., Chi, Z., Dong, L., & Wei, F. (2024). Optimizing prompts for text-to-image generation. *Advances in Neural Information Processing Systems*, *36*.

Ho, J., Chan, W., Saharia, C., Whang, J., Gao, R., Gritsenko, A., Kingma, D. P., Poole, B., Norouzi, M., N., Fleet, D., & Salimans, T. (2022). Imagen video: High definition video generation with diffusion models. *arXiv preprint arXiv.* https://doi.org/10.48550/arXiv.2210.02303

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in neural information processing systems*, *33*, 6840-6851. https://doi.org/10.48550/arXiv.2006.11239

Horvath, A. S., & Pouliou, P. (2024). AI for conceptual architecture: Reflections on designing with text-to-text, text-to-image, and image-to-image generators. *Frontiers of Architectural Research, 13*(3), 593-612. https://doi.org/10.1016/j.foar.2024.02.006

Huang, J., Johanes, M., Kim, F. C., Doumpioti, C., & Holz, G. C. (2021). On GANs, NLP and architecture: combining human and machine intelligences for the generation and evaluation of meaningful designs. *Technology| Architecture+ Design*, *5*(2), 207-224. https://doi.org/10.1080/24751448.2021.1967060

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, *55*(9), 1-35. https://doi.org/10.1145/3560815

Liu, V., & Chilton, L. B. (2022). Design guidelines for prompt engineering text-to-image generative models. In *Proceedings of the 2022 CHI conference on human factors in computing systems* (pp. 1-23.

Luo, Z., & Huang, W. (2022). FloorplanGAN: Vector residential floorplan adversarial generation. *Automation in Construction*, *142*, 104470. https://doi.org/10.1016/j.autcon.2022.104470

Nauata, N., Chang, K. H., Cheng, C. Y., Mori, G., & Furukawa, Y. (2020). House-gan: Relational generative adversarial networks for graph-constrained house layout generation. In *Computer Vision–ECCV 2020: 16th European Conference*, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16 (pp. 162-177). Springer International Publishing.

Oppenlaender, J. (2023). A taxonomy of prompt modifiers for text-to-image generation. *Behaviour & Information Technology*, 1-14. https://doi.org/10.1080/0144929x.2023.2286532

Oppenlaender, J., Linder, R., & Silvennoinen, J. (2023). Prompting ai art: An investigation into the creative skill of prompt engineering. *arXiv preprint arXiv.* https://doi.org/10.48550/arXiv.2303.13534

Paananen, V., Oppenlaender, J., & Visuri, A. (2023). Using text-to-image generation for architectural design ideation. *International Journal of Architectural Computing*, 14780771231222783.

Pavlichenko, N., & Ustalov, D. (2023). Best prompts for text-to-image models and how to find them. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 2067-2071). https://doi.org/10.48550/arXiv.2209.11711

Phenaki. (2023). Phenaki: A model for generating videos from text, with prompts that can change over time, and videos that can be as long as multiple minutes. https://phenaki.video/

Psarra, S. (2009). *Architecture and narrative: the formation of space and cultural meaning*. Routledge.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S. & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In *International conference on machine learning* (pp. 8748-8763. PMLR. https://doi.org/10.48550/arXiv.2103.00020

Rand, A. (2015). *The Fountainhead*. Penguin Publishing Group.

Ross, S. M., & Morrison, G. R. (2013). Experimental research methods. In *Handbook of research on educational communications and technology* (pp. 1007-1029. Routledge.

Runway ML. (2024). *Runway ML Web Site*. https://runwayml.com/

Sahoo, P., Meharia, P., Ghosh, A., Saha, S., Jain, V., & Chadha, A. (2024). Unveiling Hallucination in Text, Image, Video, and Audio Foundation Models: A Comprehensive Survey. *arXiv preprint arXiv.* https://arxiv.org/pdf/2405.09589v1

Saussure, F. M. (2011). *Course in general linguistics*. Columbia University Press.

Smith, E. (2022). *The traveler's guide to the latent space*. https://sweet-hall-e72.notion.site/A-Traveler-s-Guide-to-the-Latent-Space-85efba7e5e6a40e5bd3cae980f30235f

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., & Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of Machine Learning Research*, 37:2256-2265. https://proceedings.mlr.press/v37/sohl-dickstein15.html.

Steinfeld, K. (2023). Clever little tricks: a socio-technical history of text-to-image generative models. *International Journal of Architectural Computing*, *21*(2), 211-241.

Sun, J., Wu, W., Liu, L., Min, W., Zhang, G., & Zheng, L. (2022). WallPlan: synthesizing floorplans by learning to generate wall graphs. *ACM Transactions on Graphics (TOG)*, *41*(4), 1-14.

Whang, S. E., Roh, Y., Song, H., & Lee, J. G. (2023). Data collection and quality challenges in deep learning: A data-centric ai perspective. *The VLDB Journal*, *32*(4), 791-813.

Wu, A. N., Stouffs, R., & Biljecki, F. (2022). Generative adversarial networks in the built environment: A comprehensive review of the application of GANs across data types and scales. *Building and Environment*, 109477.

Zhao, L., Mo, Q., Lin, S., Wang, Z., Zuo, Z., Chen, H., Xing, W. & Lu, D. (2020). Uctgan: Diverse image inpainting based on unsupervised cross-space translation. In *Proceedings of The IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5741-5750).