REPUBLIC OF TÜRKİYE

ALTINBAŞ UNIVERSITY

Institute of Graduate Studies

Architecture

ASSESSING THE UTILITY OF TEXT-TO-IMAGE TOOLS IN GENERATING VARIOUS ARCHITECTURAL REPRESENTATIONAL TECHNIQUES

Nesrine BEKHTA

Master's Thesis

Supervisor

Asst. Prof. Dr. Can UZUN

Istanbul, 2024

ASSESSING THE UTILITY OF TEXT-TO-IMAGE TOOLS IN GENERATING VARIOUS ARCHITECTURAL REPRESENTATIONAL TECHNIQUES

Nesrine BEKHTA

Architecture

Master’s Thesis

ALTINBAŞ UNIVERSITY

2024

The thesis titled ASSESSING THE UTILITY OF TEXT-TO-IMAGE TOOLS IN GENERATING VARIOUS ARCHITECTURAL REPRESENTATIONAL TECHNIQUES prepared by NESRINE BEKHTA and submitted on 11.06.2024 has been **accepted unanimously** for the degree of Master of Science in Architecture.

Asst. Prof. Dr. Can UZUN

Supervisor

Thesis Defense Committee Members:

Asst. Prof. Dr. Can UZUN	Department of Architecture, Altınbaş University	
Asst. Prof. Dr. Erinç ONBAY	Department of Interior Architecture and Environmental Design, Altınbaş University	
Assoc. Prof. Dr Orkan Zeynel GÜZELCİ	Department of Interior Architecture, İstanbul Technical University	

I hereby declare that this thesis meets all format and submission requirements for a Master's thesis.

I hereby declare that all information/data presented in this graduation project has been obtained in full accordance with academic rules and ethical conduct. I also declare all unoriginal materials and conclusions have been cited in the text and all references mentioned in the Reference List have been cited in the text, and vice versa as required by the abovementioned rules and conduct.

Nesrine BEKHTA

DEDICATION

It would be my sincere pleasure to dedicate this thesis to those who helped me complete it, to those who were with me during my thesis preparation and writing process and stood by me emotionally and financially. I dedicate this work to my family, my parents, and my younger sister, who have supported and encouraged me all this time. I also extend my appreciation to my friends who were behind me in every step, as well as to all my instructors, and a special thanks go to my respected advisor, Asst. Prof.Dr. Can Uzun, who helped me a lot throughout my master's journey.

Finally, I hope to send gratitude to every person who was and is still supporting me, and I hope I can be up to your expectations as well as the expectations of my younger self, who dreamed of these achievements. God willing, I hope to achieve even more success.

ABSTRACT

ASSESSING THE UTILITY OF TEXT-TO-IMAGE TOOLS IN GENERATING VARIOUS ARCHITECTURAL REPRESENTATIONAL TECHNIQUES

BEKHTA, Nesrine

M.Sc., Department of Architecture, Altınbaş University

Supervisor: Asst. Prof. Dr. Can UZUN

Date: June/2024

Pages: 84

Over time, the field of architecture has seen and passed by various technologies, from the simple usage of computers to the generative AI era that we live nowadays. Text-to-image AI is a current state-of-the-art technology, that is revolutionizing today's life, by being used in many domains and for various purposes. Architecture is one of the realms that has been adopting this technology recently. This thesis aims to explore the use of text-to-image tools in the field, by investigating their potential for shaping one of the basics of architectural design, which is architectural representational techniques, through a comparative study based on rating, by generating various architectural representational techniques using five different text-to-image tools available online, and presented to architects raters to rate the different generated images by these tools for the different representational techniques. The results of this study show the different aspects of text-based image-generation tools and how they can handle the generation of the representational techniques differently and propose the use of a multi-tool approach for better results while considering the variety of the generated images, which provides valuable insight into the use of these technologies in a more efficient way and can also shape the basics for future studies for deeper discoveries.

Keywords: Text-to-image, Generative AI, Architectural Representational Techniques, Generative Design, AI In Architecture, Diffusion Models For Architecture.

ÖZET

ÇEŞİTLİ MİMARİ TEMSİL TEKNİKLERİNİN ÜRETİLMESİNDE METİNDEN GÖRÜNTÜYE ARAÇLARIN YARARLILIĞININ DEĞERLENDİRİLMESİ

BEKHTA, Nesrine

Yüksek Lisans, Mimarlık Ana Bilim Dalı, Altınbaş Üniversitesi

Danışman: Dr. Öğr. Üyesi Can UZUN

Tarih: Haziran/2024

Sayfa: 84

Zamanla mimarlık alanı, bilgisayarların basit kullanımından, günümüzde yaşadığımız üretken yapay zeka çağına kadar çeşitli teknolojileri gördü ve geçti. Metinden görüntüye yapay zeka, birçok alanda ve çeşitli amaçlarla kullanılarak günümüz yaşamında devrim yaratan güncel bir teknolojidir. Mimarlık da bu teknolojiyi son dönemde benimseyen alanlardan biri. Bu tez, metinden görüntüye araçlarının bu alanda kullanımını, mimari tasarımın temellerinden biri olan mimari temsil tekniklerini şekillendirme potansiyellerini, derecelendirmeye dayalı karşılaştırmalı bir çalışma yoluyla, çeşitli mimari yapılar üreterek araştırmayı amaçlamaktadır. Çevrimiçi olarak mevcut olan beş farklı metinden resme aracını kullanan temsil teknikleri ve bu araçlar tarafından farklı temsil teknikleri için oluşturulan farklı görüntüleri derecelendirmek üzere mimar değerlendiricilerine sunulmuştur. Bu çalışmanın sonuçları, metin tabanlı görüntü oluşturma araçlarının farklı yönlerini ve temsil tekniklerinin oluşturulmasını nasıl farklı şekilde ele alabileceklerini göstermekte ve

oluşturulan görüntülerin çeşitliliğini göz önünde bulundurarak daha iyi sonuçlar için çoklu araç yaklaşımının kullanılmasını önermektedir. Bu teknolojilerin daha verimli bir şekilde kullanılmasına ilişkin değerli bilgiler sağlayan ve aynı zamanda daha derin keşifler için gelecekteki çalışmaların temellerini şekillendirebilecek olan.

Anahtar Kelimeler: Metinden Görüntüye, Üretken Yapay Zeka, Mimari Temsil Teknikleri, Üretken Tasarım, Mimarlıkta Yapay Zeka, Mimarlık İçin Difüzyon Modelleri.

TABLE OF CONTENTS

Pages

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

LIST OF CHARTS

ABBREVIATIONS

AI	:	Artificial Intelligence
ML	:	Machine Learning
DL	:	Deep Learning
GAN	:	Generative Adversarial Networks
VAE	:	Variational Autoencoders

1. INTRODUCTION

While observing an image, we can swiftly understand its contents and the meaning of its objects is decipherable without the need for lengthy descriptions. This phenomenon holds particularly true in the realm of architecture, Architects face intricate projects and designs that can be visually compelling but solely described through text, that is the importance of visual representations in the field of architecture.

From generative design, which includes generative systems (e.g., L-systems, cellular automata), to generative AI (artificial intelligence) that can produce novel and meaningful content based on training data (Feuerriegel et al., 2024), the architecture field's commitment embraces various ways in the pursuit of creating pleasing architectural content.

In the field of architecture, artificial intelligence, a technology that mimics human intelligence, is being used more frequently as a flexible tool and has been used and included in many tasks such as the creation of building layouts, the optimization of building performance, and also building simulation and using computer programs that work using AI algorithms (Yildirim, 2022a).

Text-to-image synthesis belongs to the field of generative AI, which in turn is a part of artificial intelligence that deals with the distinct properties between textual inputs and vision as well as proficiently deciphering the jumbled written descriptions (H. Zhang et al., 2021). The field of architecture has embraced this technology, and various studies have adapted it in various manners due to its potential to facilitate the conceptualization of ideas using natural language, which alters the model of communicating ideas (Paananen et al., 2023).

Architectural representational techniques are crucial in the design workflow and they ensure the development, communication, and visualization of design ideas. They are visual systems that can be thought of as 'languages', known as spatial representation systems, and serve as mediums through which information passes (Alberto, 1982), and Allen (2009) states that they are not neutral and leave their trace on the final outcome. According to de la Fuente Suárez (2016), any architectural representation involves the process of the representation of experience and the process of experience of the representation, and any representation has

two layers, the ‘what’ which is the object, and the ‘how’ which is the representational medium.

1.1 AIM OF THE STUDY

This thesis aims to showcase the contribution of artificial intelligence (AI) in the field of architecture, specifically assessing the usefulness of different text-to-image tools that are part of generative AI in the task of generating different architectural representational techniques that are bubble diagram, conceptual sketches, floor plan drawings, cross-section drawings, elevation drawing, exploded view diagram, rendered 3D model and the architectural element details, by evaluation this generated contents relying on the perception of architects. By doing so, this study can help shape the basics of adopting these technological tools in the design assignment and influence further studies to explore more deeply the use of these state-of-the-art methods in architecture due to the capability of the field to keep up with the times. This thesis addresses the following questions: Do currently available text-to-image tools possess knowledge of the architectural framework and how architects are experiencing them? Specifically for the various architectural representational techniques, how good are these tools regarding this task? How do these tools process the various architectural representational techniques? Given the same text input, will the same text-to-image tool generate different images with the same consistency?

2. LITERATURE REVIEW

In order to understand the topic of using text-to-image tools and specifically assessing their utility in generating various architectural representational techniques, it is crucial to understand the background of the topic as a whole. This literature review section examines the different components that contribute to understanding the subject matter in a top-down manner starting with artificial intelligence and its subfields and the generative models in AI, and then progressing to the topic of text-to-image generation in the field of architecture.

2.1 ARTIFICIAL INTELLIGENCE

Since more than six decades ago, artificial intelligence has evolved with different techniques and subfields ranging from rule-based systems to Machine learning and many others (Wirz & Gallo, 2020). Artificial intelligence is processing Several tasks that are intended for humans to solve, through several approaches instanced by enhancing, simulating, and increasing human intelligence when mimicking the activity of a single network was the attempt of the earliest AI models with a simple input-output function that got sophisticated over decades (Forghani, 2020). Artificial intelligence represents two dimensions, the dimension related to thinking and the dimension related to behaving (Delipetrev et al., 2020). The pioneering idea of artificial intelligence goes back to ancient Greek mythology with the idea of humanoid robots (Mijwil, 2015) in the journalism of that era when mechanical devices acted with seemingly intelligent behavior (Jeffery, 2022).

The roots of AI can probably be back to 1942 with the history developed by the American Science Fiction writer Isaac Asimov about a robot that follows the laws of robotics which inspired the scientists (Haenlein & Kaplan, 2019) and the term "Artificial Intelligence" first saw the light officially in 1956 when a summer research project at Dartmouth was hosted (Haenlein & Kaplan, 2019) (Delipetrev et al., 2020) (Kelleher, 2019). AI Throughout history has passed through different periods as (Figure 2.1) shows. In 1951 the first functioning AI programs were written for playing programs (Jeffery, 2022), and then in the 1970s AI saw its first winter due to several issues which resulted in a decline in AI activities, then in 1980s it was aimed to capture human knowledge and move it to a computational form in the paradigm of AI called "expert system" which saw it decline for the IT lexicon after 1990s

and it is known as the second AI winter which is followed by the spring of AI that saw the light in the 2010s by the development of deep learning (Delipetrev et al., 2020).

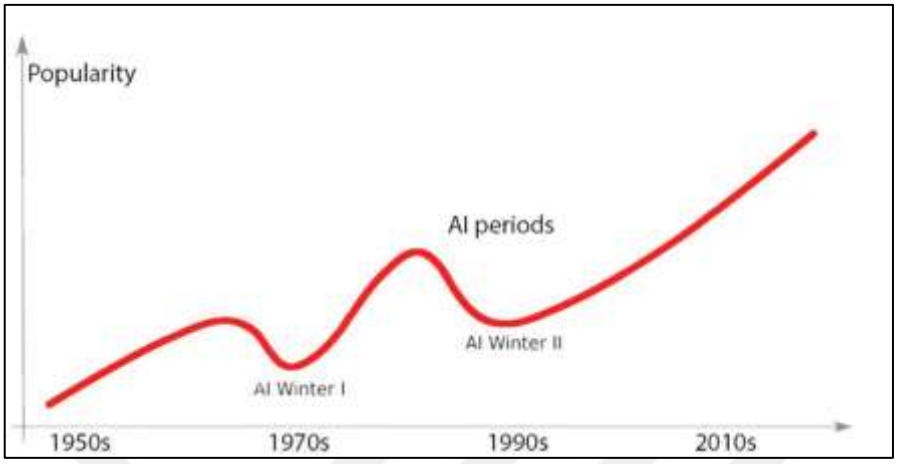


Figure 2.1: The Curve of AI Periods (Delipetrev et al., 2020).

Machine learning and deep learning are subfields of AI that use sophisticated algorithms and multi-layered neural networks to address problems (Ghosh & Thirugnanam, 2021). Machine learning is a technology that allows computers to learn from previous data and make predictions of the outcome based on these data by using mathematical models based on diverse techniques without being explicitly programmed (Shaveta, 2023) and it allows computers to learn, self-teach, and improve as they use huge datasets(Hua, 2022). Machine learning can be classified into three categories supervised, unsupervised, and reinforcement learning (Hua, 2022) (Shaveta, 2023) (Chaudhary & Vasuja, 2019). Supervised learning is for comparing the expected output with the actual one and addressing the errors to be solved to reach the expected outcome, this kind of learning can be classified as regression and classification, while unsupervised learning involves auto-learning by computers based on adopting and discovering from input patterns and it is classified into association and clustering (Chaudhary & Vasuja, 2019) as it is shown in (Figure 2.2), while in reinforcement learning, the agent is rewarded for its right action and incurs a penalty for each incorrect one, thus it tries to perform the best to be rewarded with the most points (Shaveta, 2023).

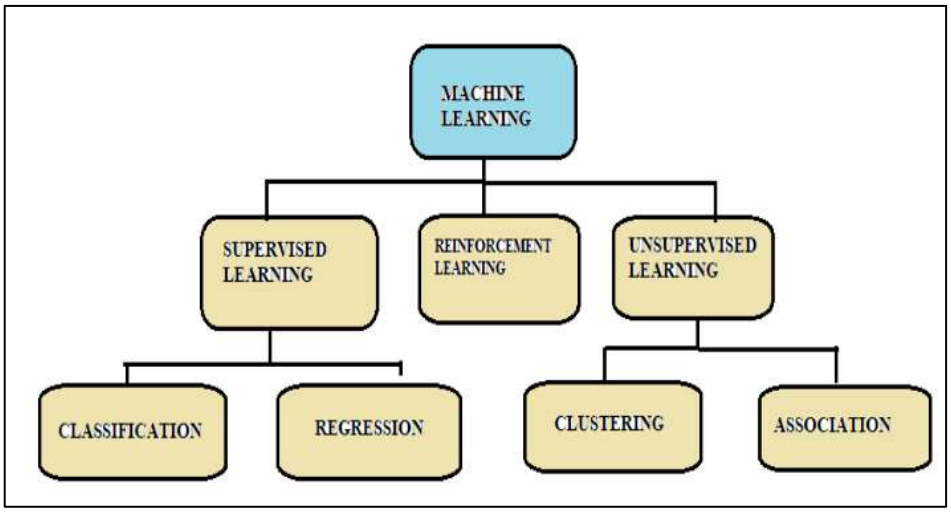


Figure 2.2: Machine Learning Categorization Graph (Chaudhary & Vasuja, 2019).

Deep learning is a subfield of artificial intelligence, specifically Machine learning, that is effective in handling large and complex datasets based on developing large neural network models (Kelleher, 2019), and deep learning algorithms algorithms have achieved noticeable success in various fields instanced by natural language processing and computer vision (Olaoye & Potter, 2024). Deep learning mimics the activity in layers of neurons found in the part that represents 80% of the brain where thinking takes place and acquires knowledge about the different levels and hierarchies of abstraction and representation in order to recognize the patterns of data that come from various sources such as images and texts (Bhardwaj et al., 2018).

2.2 GENERATIVE MODELS IN AI

To entirely understand the potential of text-to-image generative models, it is crucial to understand their scope as well as their background, for this reason, Generative AI, GANs, VAE, Diffusion Models, and image synthesis, will be discussed in trying to gain enough knowledge about the field, and then text-to-image models will be presented.

2.2.1 Generative AI

The term generative AI describes the computational techniques that can produce novel and meaningful content based on training data (Feuerriegel et al., 2024), its rapid growth has occurred thanks to the advances in deep learning techniques, as well as the accessibility of large-scale datasets (Bandi et al., 2023), (Figure 2.3) shows generative AI within the

hierarchy of artificial intelligence and basic definitions of the used terms. Generative AI can be used across various modalities including video, image, audio, text, code, as well as other contexts (Banh & Strobel, 2023). Cao et al. (2023) state that the creation of digital content is done through AI models and it can be unimodal if they follow instructions in the same format as the generated content or multimodal that accepts cross-modal instruction as seen in (Figure 2.4).

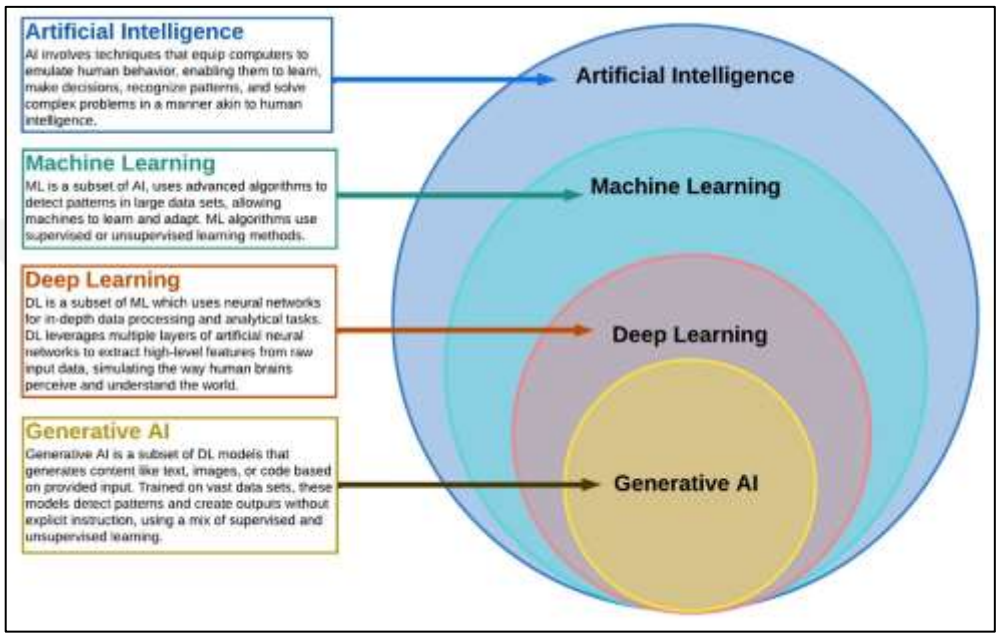


Figure 2.3: Generative AI Within The Hierarchy of Artificial Intelligence (Zhuhadar, 2023).

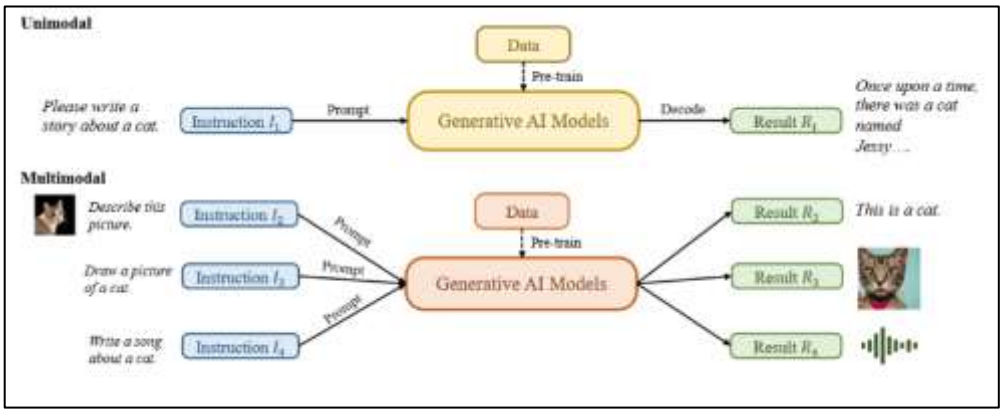


Figure 2.4: Unimodal and Multimodal Models of GAI (Cao et al., 2023).

An investigation of the key techniques of image synthesis generative AI that are essential components including Generative Adversarial Networks, Variational Autoencoders, and

diffusion models and their differences, and understanding their concepts helps to build the knowledge of text-to-image generation.

2.2.1.1 Generative adversarial networks GANs

Generative Adversarial Networks (GANs) were first introduced by Goodfellow et al. (2014), and they consist of two different networks, a generator network G that is trained to deceive the discriminator D which is trained to determine if a particular image is fake, or real. The objective of discriminator D is to correctly classify the data source, and the objective of generator G is to enhance the performance of generated data, thus, If the input comes from the generator G(z) that uses noise z, the discriminator D should classify it as false and label it as 0 if the input comes from real data then the discriminator D should classify it as true and label it as 1 (K. Wang et al., 2017), the framework of GANs in the case of image processing and an exampl of their training are shown in (Figure 2.5). GANs have advantages over other generative algorithms such as the few constraints on the design of the generator, and the capability of GANs to parallelize the generation as well as their great success achieved in the realm of computer vision and image processing (Gui et al., 2023), their fame may also be perceived through their various existing variants as (Figure 2.6) demonstrates.

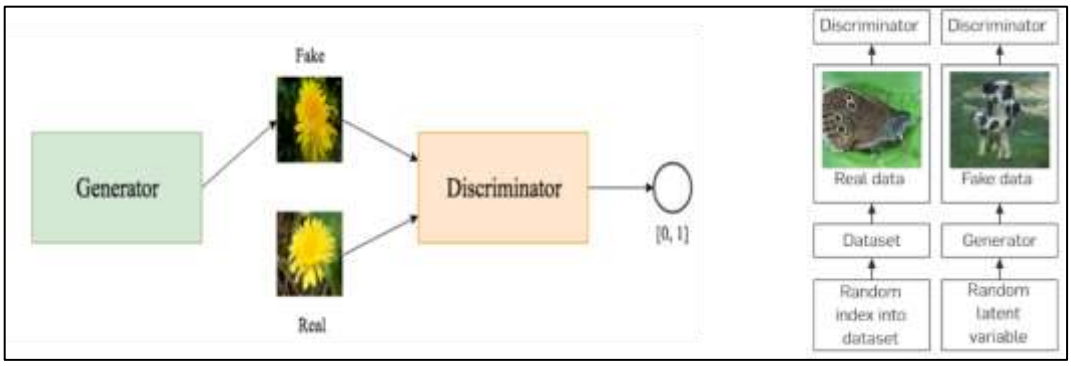


Figure 2.5: Left: The Framework of GANs in the Example of Image Processing. Right: The Training of Gans That Necessitates Training Both, The Generator and The Discriminator Network GANs (Bodnar, 2018) (Goodfellow Et Al., 2020).

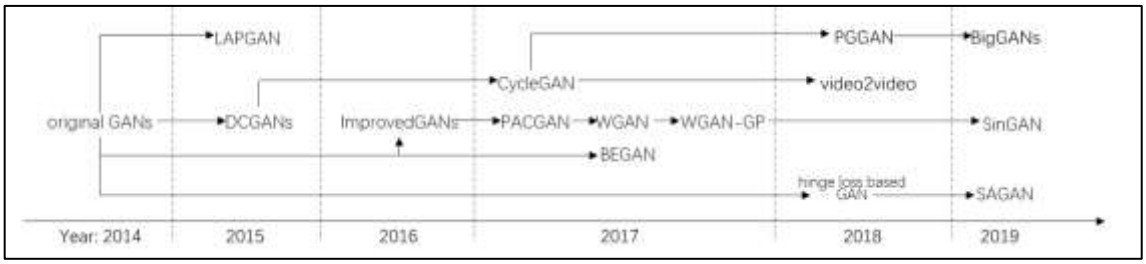


Figure 2.6: The Timeline of GANs Variants (Gui et al., 2023).

2.2.1.2 Variational autoencoders VAE

Variational Autoencoders were first introduced by Kingma & Welling (2013), and are generation models that leverage a latent space that achieved an extensive application in data generation (J. Liu, 2023). According to Baraheem et al. (2023), variational autoencoders are based on a decoder and an encoder and their training seeks to decrease the amount of error that takes place during reconstruction between the original data and the data after encoding and decoding, when the input data is converted into the distribution that gets decoded into the original input data (D. Wang et al., 2023). The general framework of VAE can be seen in (Figure 2.7).

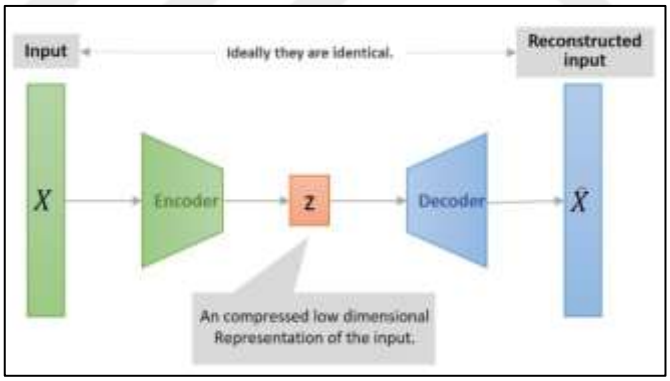


Figure 2.7: The General Framework of VAE (F. Liu & Qian, 2022).

2.2.1.3 Diffusion models

Recently, one of the widely used models for text-to-image generation is diffusion models, inspired by the original work of Sohl-Dickstein et al. (2015), their core concept is influenced by non-equilibrium statistical physics and helps to gradually and systematically break down the structure in a data distribution through an iterative forward diffusion process and then learn to reverse this process and rebuild the patterns. They are known as “diffusion” models

due to the noise that is being diffused for non-noisy data, and then the model learns how to reverse the process by denoising the data. Diffusion models belong to probabilistic generative models that progressively generate samples by introducing noise to data and then learn to reverse this process (Yang et al., 2023; Croitoru et al., 2023). The Gaussian noise is added progressively to the input until it becomes completely noisy, and then the noise is extracted gradually, which is called the forward diffusion process and the backward diffusion process, respectively (Croitoru et al., 2023; Fishman et al., 2024). (Figure 2.8) shows the idea behind diffusion models. Rombach et al. (2022) state that the advantages of diffusion models include avoiding issues like mode-collapse and training instabilities, and effectively modeling complex distributions of natural images without requiring a huge number of parameters.

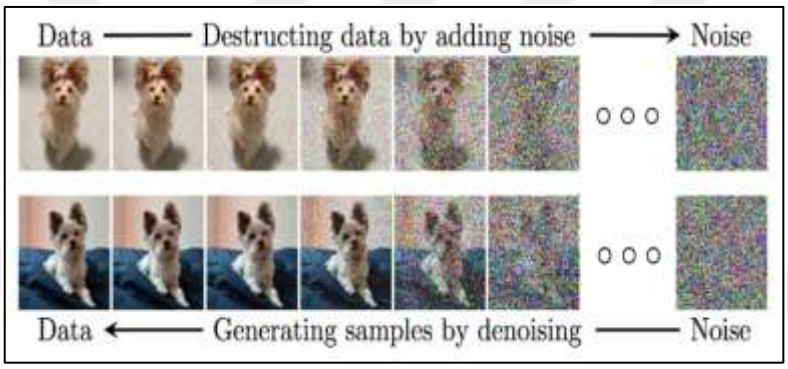


Figure 2.8: The Idea Behind The Diffusion Models, Modified From Source (Yang et al., 2023).

2.2.2 Image Synthesis

One possible application of generative AI is image synthesis, worth noting that generating images using computers is not a novel concept and dates years ago when the idea of computer-generated image content gained momentum in the 1970s (Tsirikoglou et al., 2020). Deep learning has a crucial role in the image synthesis domain, and Tsirikoglou et al. (2020) state that since its introduction to generative modeling, the emergence of image generation methods that do not belong to traditional image synthesis pipelines has appeared (e.g. GANs and VAE), and it achieved significant success in image processing and it is the dominant method in this field nowadays (Jiao & Zhao, 2019). Baraheem et al. (2023) define image synthesis as a significant issue in the computer vision field when attempts have been made by the research community to generate high-resolution images, some methods used to

generate images can be either, conditioned which means to guide the generation process and can take the form of class labels, text, or other forms (Graikos et al., 2023), or unconditioned, which means generating images without guidance, when Yang et al. (2023) state that generative models were initially developed for unconditional generation, and they subsequently closely pursued the development of conditional content, and both can be achieved using several types of generative models. (Figure 2.9) shows the varieties of image synthesis using several models and the different types of input data.

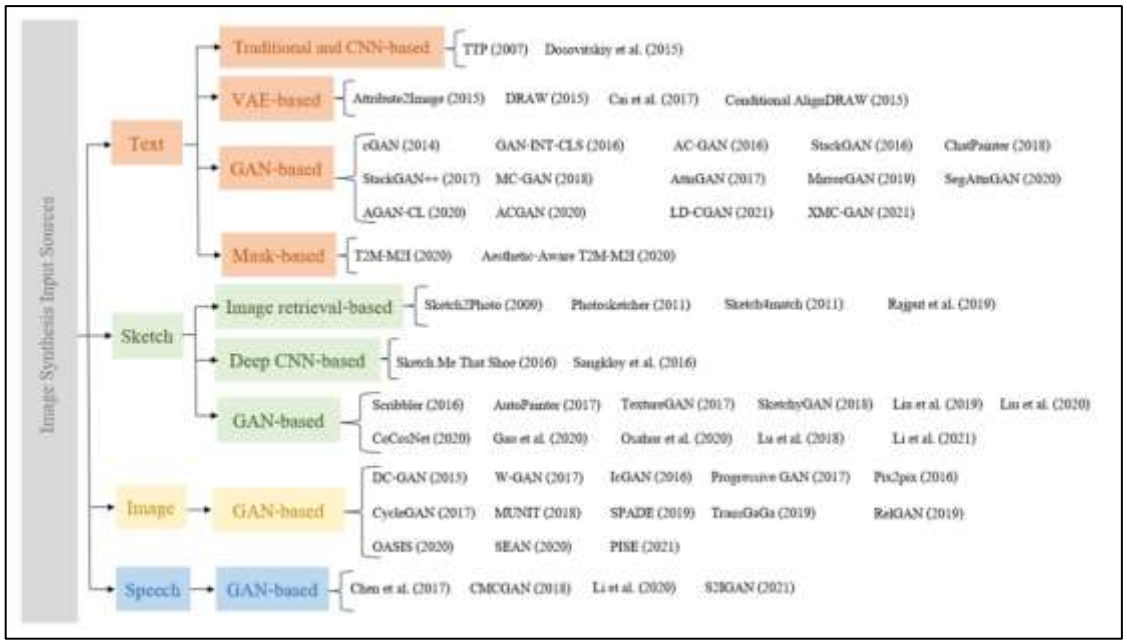


Figure 2.9: Classification Of Image Synthesis (Baraheem et al., 2023).

2.2.3 Text-To-Image Generation

Text-to-image is a multi-model task seeking to translate the textual inputs into high-quality images (Xie et al., 2023), and it involves the intersection of the domain of computer vision and natural language processing (Jin et al., 2023). The timeline of the advancements of text-to-image generation up to 2022 is shown in (Figure 2.10**)**, noting that the field is growing rapidly and many other developments are joining the field. The previously discussed generative models have also been used for text-to-image generation tasks when text input plays the role of the condition that guides the generative models. modern machine-learning approaches for text-to-image generation began with the research conducted by Mansimov et al. (2015) as discussed by Ramesh et al. (2021) and Zhang et al. (2023).

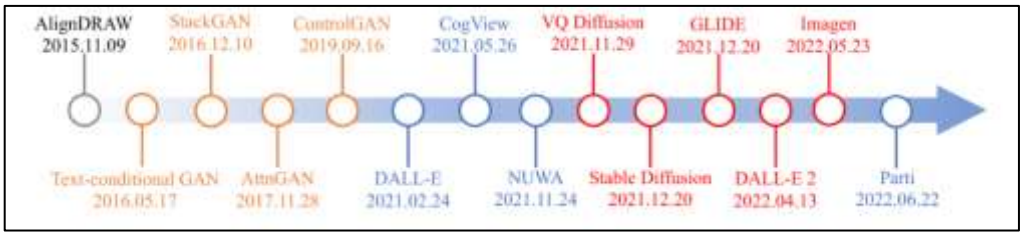


Figure 2.10: The timeline of works on text-to-image generation tasks (C. Zhang et al., 2023).

Mansimov et al. (2015) attempted to develop a model (Figure 2.11) that can be used in practical situations since images are usually accompanied by unstructured textual explanations, this new model is adapted from Deep Recurrent Attention Writer (DRAW) (Gregor et al., 2015), that belongs to the family of variational autoencoders VAEs, to generate images while taking into account textual descriptions using a Bidirectional recurrent neural network (RNN) as a language model. Additionally, when it creates an image, it dynamically focuses on the pertinent phrases in the accompanying textual description. In this model, images are portrayed as a series of patches drawn on a Canvas, and captions are depicted as a series of consecutive words, it has been trained and evaluated using the Microsoft COCO dataset (T.-Y. Lin et al., 2014). However, as a result, it often produces samples that are slightly unfocused, even after trying to augment the model using a technique to enhance sharpness afterward.

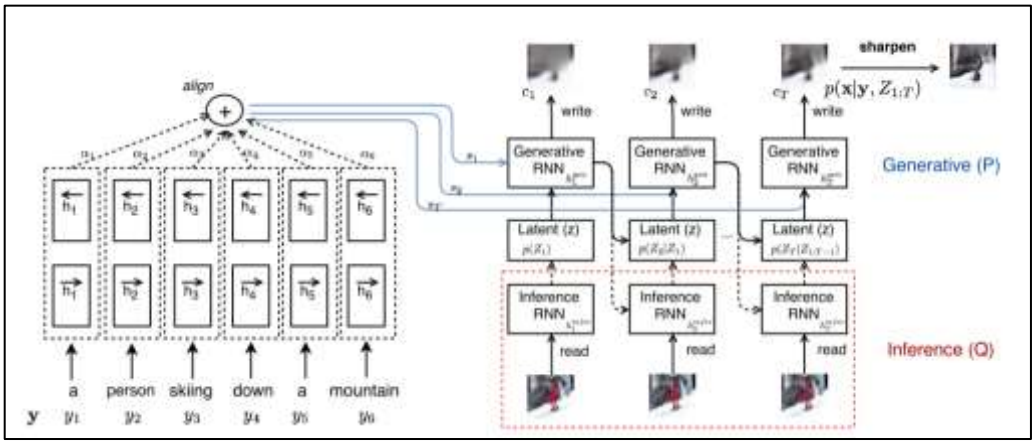


Figure 2.11: Right: AlignDRAW Model for Generating Images by Getting an Alignment Between the Generating Canvas and the Input Captions. Left: Caption Encoding in Two Directions Using Bidirectional RNN. Right; Image Generation (Mansimov et al., 2015).

GAN algorithms are also used for text-to-image synthesis**,** and some of its variants have been used for this task. Zhou et al. (2021) categorize the improvements in generating images from text using GANs into three groups, and GANs variants belong to these categories of improvements, (Figure 2.12) shows the use of GANs in the task of text-to-image generation as well as the improvements categories. For instance, Xu et al. (2018) proposed an Attentional Generative Adversarial Network (AttnGAN), that generates images by first paying attention to the different parts of the textual description, after that the matching loss between the image and text is counted to train the generation. Another example can be seen in the work of Zhang et al. (2017) when Stacked Generative Adversarial Networks (StackGAN), the proposed approach generates images from text passing through stages I, when the object is outlined first based on the text input and then stage II in which it gets corrected with the addition of further details.

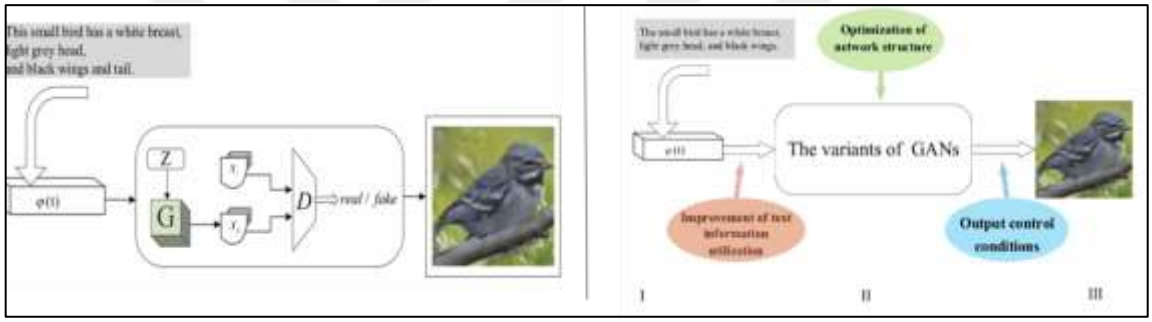


Figure 2.12: Left: General Text-to-image Synthesis Using GANs, When the Text is Embedded First Into Vectors, Then the Generator G Creates Fake Images, When Xg Represents the Image Generator Receiving the Text Features Φ(T) And The Noise Vector Z, And Xr Represents the Distribution Of Real Samples Used By the Discriminator. Right: The Improvements for the Three Steps of Text-to-image Generation Using GANs (Zhou et al., 2021).

Other more advanced approaches have appeared, instanced by the work of Ramesh et al. (2021) when a transformer-based (Vaswani et al., 2017) method is used, a straightforward method has been suggested using an autoregressive transformer that has been trained in a two-stage procedure, compress each large image into a grid of smaller parts, and train an autoregressive transformer to understand how the text and image parts are related, respectively, when applied on a large scale. Gafni et al. (2022) proposed a new text-to-image method using an autoregressive transformer, the method involves integrating a straightforward control mechanism that complements text using scenes, enhances the

tokenization process using some components, and adjusts transformer guidance without the need for classifiers, additionally, this method provides also novel capabilities such as dealing with atypical text instruction and scene editing.

2.2.3.1 State-of-the-art text-to-image generators

Several models that are used to generate images from textual descriptions have been developed in recent years, and many of these approaches have been presented in research papers and articles to advocate and elucidate their methods. It is worth noting that not all the generators presented in research papers are available to the public or they do not have an available online interface, the different models are presented here to highlight the differences that may exist between them even though some of them used the same model.

GLIDE is a text-to-image model that was introduced by Nichol et al. (2022) when investigating the application of diffusion models in text-based image synthesis with the incorporation of a text encoder to make it dependent on descriptions in natural language when the text is first encoded into tokens, and then fed into a transformer model. The process was done by using two different guidance strategies CLIP (Radford et al., 2021), and classifier-free guidance.

DALL-E 2 is one of the most famous text-to-image models. Ramesh et al., (2022) proposed a combination of diffusion models commonly used in generation tasks with CLIP embeddings that its latent space has the potential to semantically modify images by shifting towards any encoded text direction, this combination starts by first training a diffusion decoder to reverse the CLIP image encoder while this diffusion decoder can then generate multiple images from a given image embedding. An overview of DALL-2 can be seen in (Figure 2.13). DALLE-2 has been compared with another model which is Imagen, Saharia et al. (2022) found that Imagen performs better than other simultaneous efforts of DALL-E 2, and Imagen was presented as a text-to-image diffusion model that merges the advanced high-quality diffusion models with the capabilities of large transformer language model and this combination results in a high level of photorealism and a profound level of language understanding in text-to-image synthesis has been trained on text embeddings from large language model instead of image-text data and it uses text-conditional super-resolution diffusion models to upsample the image, (Figure 2.14) shows the workflow of Imagen.

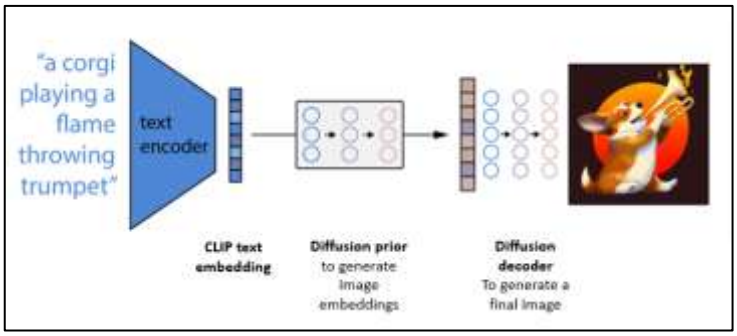


Figure 2.13: A High-level View of the General Structure of DALL-E 2, Adapted From Source (Ramesh et al., 2022) and (O'Connor, 2023).

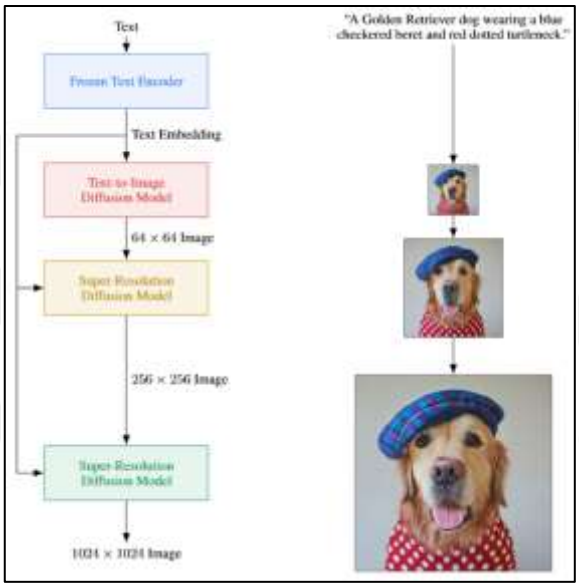


Figure 2.14: Workflow of Imagen; The text Is Encoded to Embeddings, Which Are Mapped Into a Low-resolution Image Using a Diffusion Model, and Then the Image Gets Upsampled Using Super-resolution Diffusion Models (Saharia et al., 2022).

Stable diffusion is also another text-to-image generator model, in which images are denoised to get a high-resolution image that matches the textual input (Lee et al., 2023), with the presence of several components that form the whole model (Alammar, 2022). The model can efficiently operate on consumer-grade GPUs, allowing it to generate high-quality images at a high resolution in a matter of seconds, and it incorporates valuable insights from conditional diffusion models and is built upon the work of latent diffusion models (Stable Diffusion Launch Announcement, 2022). In the work of Rombach et al. (2022a), the departure to latent space to synthesize high-resolution images is discussed, by first analyzing the trained diffusion models in pixel space and then undergoing a two-stage learning, the

stage of perceptual compression that eliminates high-frequency details while keeping minimal semantic variation, and then the generative model learns the semantic and conceptual composition of the data, both with cross attention conditioning mechanism to allow the multi-modal training, help to improve the training and sampling efficiency of denoising diffusion models greatly while maintaining their quality across various conditional image synthesis tasks (Figure 2.15).

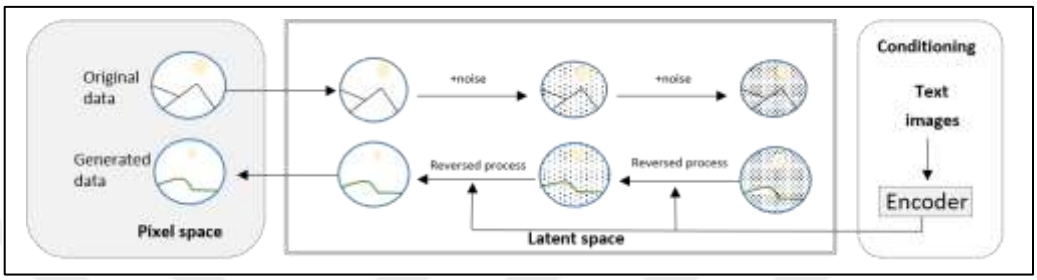


Figure 2.15: The Conditioning of the Latent Space, Adapted From Source (Rombach et al., 2022a).

DALL-E 3 is one of the newest text-to-image generators by open AI and is more improved than DALL-E 2. DALL-E 3 has been added as an image generator Constituent to ChatGPT (OpenAI, 2023). DALL-E 3 is a latent diffusion model that applies caption improvement to deal with the prompt following, by providing the captioner with captions that describe the general context of the images and significantly descriptive captions that capture more details (Betker et al., 2023). DALL-E 3 can follow the details provided in the textual prompts as well as generate uncommon scenes (K. Lin et al., 2023).

2.3 TEXT-TO-IMAGE GENERATION IN THE FIELD OF ARCHITECTURE

In advance of considering the usage of text-to-image technologies in the field of architecture, it is essential to understand their applicability in some related fields. This helps to provide a wider perspective of the topic, the capabilities of these tools, as well as their features that differ from one tool to another and across different applications.

In the realm of urban design, Yildirim (2022a) states that text-to-image tools offer advantages for the field since they can help generate the desired environment quickly instead of creating each element manually, which may be advocated by the work of Seneviratne et al. (2022) when they investigated the proficiency and biases of text-to-image methods and

their application to the built environment precisely in urban planning, by generating queries related to this environment and then evaluating these generations, as a consequence, it has been found that the model can generate high-quality images in different image styles with capturing the features of in urban design, an example is shown in (Figure 2.16) when the concept of healthy and unhealthy neighborhoods has been captured. However, a limitation in creating life-like real-world scenes with high precision has been observed. This may show that the tools can be improved to accomplish urban-specific tasks and take into consideration the elements of the urban environment.

Figure 2.16: Top: The Generated Images Represent Healthy Neighborhoods. Bottom: The Generated Images Represent Unhealthy Neighborhoods (Seneviratne et al., 2022).

Interior design is a field that benefits from generating images from text. Yildirim (2022a) states that text-to-image tools offer advantages for interior design with their ability to generate realistic images quickly with the flexibility to explore a broad range of options, and it is adopted in the field across diverse applications. For instance, Loder (2023) searched the use of AI imaging tools for interior design through a workshop that involved participants from relevant fields, finding that using Modjourney AI for designing different aspects of interior design was appreciated for its unimagined outcomes, but limited at the same time since the results were not precise, which highlight the advantages and the creative outcome of AI tools that can be improved to reach the level of precision that human designers show.

These tools can be also adapted to interior design using another approach, illustrated by the study of Chen, Shao, et al. (2023) when an interior decoration style dataset was created and used to retrain and fine-tune diffusion model by proposing a novel loss function, which guides the model to generate interior design images that satisfy the requirements, an example of these images is shown in (Figure 2.17). The role that text-to-image AI can play in concept generation was also discovered as the study of Brisco et al. (2023) shows when they searched the possibility of using modern text-to-image AI such as Midjourney and DALL-E 2 to replace human designers in the stage of generating design concept, in their study, the generated images depict a chair. As a result, it was concluded that the generated images could be used as a source of inspiration. However, they still can not replace designers in the task of concept generation. The reviewed literature shows that in architecture-related fields, efforts to adopt the advanced technologies of artificial intelligence are made. Nevertheless, many improvements can be applied to meet the specific requirements of the different disciplines.

Figure 2.17: Example of Children's Room Generated by the Fine-tuned Model (Chen, Shao, et al., 2023).

In architecture, through the examination of the literature, it can be observed that text-to-image tools have been adapted in various ways and through different lenses, enriching the application of artificial intelligence in the field. Some studies have investigated the utility of these technologies in architectural education practice, which may support the learning environment and stimulate creative thinking. To cite an instance, Yildirim (2022b) examined a design problem presented to first-year architecture students to convert written quotes into dimensional reliefs through abstraction using text-to-image tools, finding that these tools, helped students expand their perspective, think abstractly and creatively, and provide them with a supple learning environment. The study of Paananen et al. (2023) is another one conducted in the educational workflow, when a laboratory study was conducted with

architecture students to develop a concept for a culture center using text-to-image tools in the first stages of the design process, The study found that these tools can have a positive impact on supporting ideas discovering and spark creative mindset when carefully incorporating design constraints. Other studies focus on how these technologies are being adopted by the field, highlighting that there is interest in adopting them, as the study of Ploennigs & Berger (2023a) shows, when a study analyzing the available queries related to the realm of architecture in the Midjourney AI platform was done, finding that architecture use cases are broadly present and these tools can be used for idea generation, variants creation, and architectural collages. Other studies have taken another approach focusing on adapting the different tools to architectural knowledge, to show the adaptability of these latest technologies in the field. Ploennigs & Berger (2023b), researched the computational design capabilities of diffusion-based AI generators, specifically for the generation of floor plans, and also developed an approach with an enhanced semantic encoding, when according to their experiments, the fine-tuning improved the validity of the generated floor plans to 90%. Likewise, Chen, Wang, et al. (2023) proposed a new text-to-image generation AI method that works through an improved diffusion model based on textual inputs, the proposed method is specified for generating designs with specific styles and master architect quality, summarizing that this method provides diversity, the optimization of the workflow, and the usability in diverse domains which provides quality and efficiency, some examples of the resulted images can be seen in (Figure 2.18).

Figure 2.18: Example of Generated Images by the Improved Diffusion odel reRpresenting Architectural Design in the Style of Franck Ghery (Chen, Wang, et al., 2023).

Other studies highlight the usefulness of these state-of-the-art tools differently. Jaruga-Rozdolska (2022) investigated the implementations of Midjourney AI in architecture and evaluated the resulting generations including concept generation, technical drawings, and

architect-specific manner drawings, showing that artificial intelligence can save architects time due to its speed. However, evaluating the usability of the buildings is the role of the architects since the output of artificial intelligence represents only the first phase of creation. Comparably, the study of Radhakrishnan (2023) sought to identify if the strategies of AI-Art can be an obstacle to the creativity linked with architectural imagery through conducting research experiments and analyzing the literature review, finding that AI art is not a competitor in the creative process, with the need to ensure that protection of the creativity in this age of automation. The importance of the designer's capabilities was highlighted by the study conducted by Z. Zhang et al. (2023), which investigated the comparison of designs, the ones generated by artificial intelligence and the ones designed by the architect Antonio Gaudi based on some metrics criteria, suggesting that in some aspects such as authenticity the designs of Gaudi were better, and in others AI outperformed emphasizing the uniqueness of human design. Tanugraha (2023) adds another contribution to this perspective, when investigating the use of Midjourney AI in supplying ideas to the vernacular architecture, revealing that Midjourney can play the role of virtual assistant and help generate concept ideas for selecting the right material, nevertheless, its capability to implement structural logic into its designs has not yet demonstrated. Albaghajati et al. (2023) explored the potential of text-to-image tools by conducting interviews with designers who experienced already text-to-image tools within architectural practice, finding that their application varies across the different design stages, and helps to enhance creativity, imagination, and visualization notably in the early design stage.

Different from these works, this thesis study concentrates on using text-to-image tools to generate images representing architectural representational techniques to highlight their significance within the architecture workflow.

3. METHODOLOGY

3.1 CONTEXT

In the literature, different distinctions of representational techniques can be found. Lockard, (1982) distinguished between qualitative representations that convey the experience of realism of the designed environment such as perspectives, and quantitative representations such as orthographic projections. In the work of Alberto (1982), the types of representation are grouped into three categories, two-dimensional, three-dimensional, and the fourth dimension. Riahi (2017) argues that drawings have been the principal medium of representation of architecture. However, lastly, alternative visual methods such as digital media expanded the horizon of architectural representation, Shojaee et al. (2022) distinct between three methods of architectural representation that belong to different periods and are, architectural Representation as imitation, and mimesis, architecture as abstraction, and architecture as simulation. Farrelly (2008), provides an in-depth classification of architectural representational techniques that are: Sketch, scale, orthographic projection, three-dimensional images, modelling, and layout and presentation, and each category contains sub-categories. The richness of this classification presents a suitable framework for this study. Sketch, scale, orthographic projection, and modeling have been selected, to explore a wide range of techniques, as well as, the inclusion of diagrams to explore in a wider way the abilities of text-to-image tools, in addition to the importance of diagrams in communicating architectural ideas since they are an important representation tool that is used to explain ideas and can be used in different stages of the design from the early to the implementation stage (Baydogan & Karamış, 2023).

To convey different aspects of architectural design, a broad range of representational techniques has been curated. From the sketch category, conceptual sketches have been selected since the case study examines a general concept. Plans, sections, and elevations have been selected as being integral components of orthographic projection. For the scale category, and to complement the orthographic projection, detail scale drawings were used. To represent a model of architectural work and since the modelling is a technique of representation, the choice has been made to utilize rendered 3D model to cover both the modeling process and rendering effects. Finally, in terms of diagrams, two distinct types

have been utilized, the bubble diagram which is used to visualize spatial connectivity (Yong & Chibiao, 2022), and the exploded view diagram which encompasses the essence of diagram as well as the meaning of three-dimensional drawing defined by Farrelly (2008) as a visual that convincingly conveys the overall sense of a project when combined with other two-dimensional drawings. Overall, the architectural representational techniques addressed in this study are conceptual sketches, plans, sections, detail scale drawings, rendered 3D model, bubble diagram, and exploded-view diagram.

To investigate the potential of different text-to-image tools for various architectural representational techniques, the case study of a two-floor house project has been selected. This selection is made based on the universality of this type of building, since it is one of the most widespread buildings that can be found anywhere, as the survey of this study is destinated to architects all over the world, and it is one of the fundamental building types that all architects, have experience with, at least, during their education, which allows discovering how text-to-image tools can generate different representational techniques for various architectural aspects that can be found in other types of buildings. The two-floor selection has been made in trying to guide the text-to-image tools more specifically to explore different spatial interactions between the floors like staircases in this study.

3.2 DATA CREATION

3.2.1 Prompt Design

A prompt is defined as an input of a natural language that the model receives and the output is an image in the case of text-to-image generation (Pavlichenko & Ustalov, 2023). Since text prompts are open-ended and free-form, it means that many ways to generate images are available, which might make generating an image become a method of exhaustive trial and error (V. Liu & Chilton, 2023). Thus, the textual input prompts must be provided in a specific format, which belongs to the research known as prompt engineering (Oppenlaender, 2023), in which Oppenlaender et al. (2023) defines 'skill in prompt engineering' as effectively using language and prior knowledge to create prompts that describe the desired outputs. Templates and guidelines of these text prompts are used to help in writing the prompts appropriately to keep the same design for all the prompts and avoid misunderstandings and biases. Recent studies have provided some templates or prompt designs that help to create and modify the

textual prompts to generate visually appealing images. Some templates are available online to help the text-to-image AI users community write their prompts. parsons (2022) developed a prompt book that guides the creation of textual prompts to use DALL-E 2 by providing different keywords for different types of images. Additionally, the work of Smith (2022) introduced the template: [Medium] [Subject] [Artist(s)] [Details] [Image repository support] that can be applied to CLIP Guided Diffusion notebooks, CLIP (Radford et al., 2021), is a multi-model for learning representations of text and image collaboratively (Bianchi et al., 2021). Moreover, some studies and research explored different ways that help formulate textual prompts in a general way. For example, the work of Pavlichenko & Ustalov (2023) provides the prompt format as follows: [kw1, . . . , kwm−1] [description] [kwm, . . . , kwn] when a prompt contains the description which is the main subject and the keywords that describe this subject and take place before and after the description, as well as the study of Oppenlaender (2023) that provides an academic understanding of text-to-image generation by providing a taxonomy of different types of text modifiers when a list of prompt modifiers was compiled, these modifiers help to direct the text-to-image system and thus modify the resulting images as seen in (Figure 3.1).

Modifier	Description
Subject term	Denotes the subject
Style modifier	Indicates an artistic style
Image prompt	Indicates a style or subject via an image
Quality booster	A term intended to improve the quality of the image
Repeating term	Repetition of subject terms or style terms with the intention of strengthening this subject or style
Magic term	A term that is semantically different from the rest of the prompt with the intention to produce surprising results

Figure 3.1: The list of Prompt Modifiers (Oppenlaender, 2023).

For this thesis study, to create the desired textual prompts, the architectural context has been taken into consideration, without mentioning any specific architect or any specific architectural period style to avoid any kind of bias. The taxonomy of Oppenlaender (2023) is used by selecting two key modifiers, [subject term] which is subject (e.g. floor plan, conceptual), and [style modifier], in this case, is the type of representations that can be drawing, sketch, model, and diagram, and also keywords representing extra information are

used as Pavlichenko & Ustalov (2023) define, and they can be placed before and after the description. In this study case they are placed last, and are ‘two-floor’ and ‘house project’. Also to provide more context understanding, another keyword is placed first in the cases where emphasizing the architectural context is necessary to highlight it and avoid any ambiguity. Such as ‘Architectural’ for the bubble diagram and the exploded view diagram, ‘rendered’ for the 3D model, and ‘technical’ ‘detailed’ for the element details. Thus, these elements represent the main idea of the textual prompt since the subject and style keywords are the most contributing words more than the connecting words (V. Liu & Chilton, 2023). Singular and plural forms of the subject terms are leveraged to guide the tools, as well as assess the ability of the tools to understand the grammatical structure. Therefore, in this study the general prompt format is ‘keyword (s)’ [subject term] [style modifier] of a ‘keyword’ ‘keyword’ ‘keyword’. Accordingly, the ‘keywords’ [subject term] [style modifier] of a two-floor house project. The prompts for each representational technique are shown in (Figure 3.2).

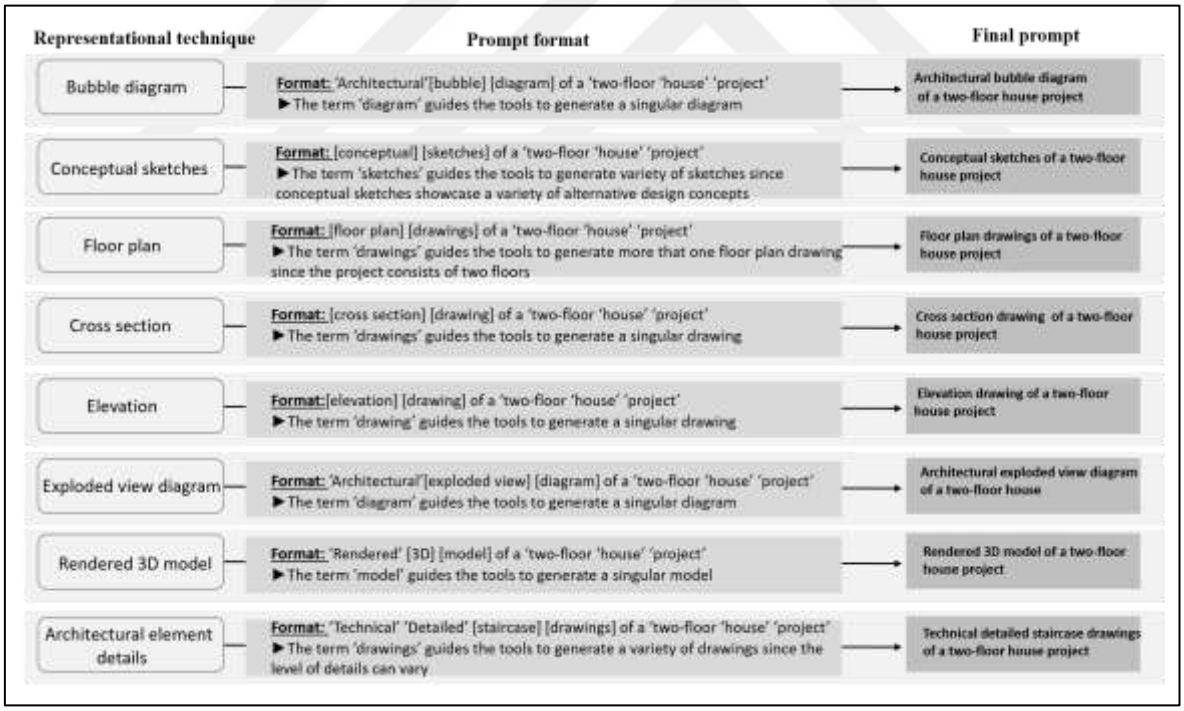


Figure 3.2: The Representational Techniques, Prompt Format, and Final Prompt (Created by the Author).

3.2.2 Images Generation: Tools And Steps

Five available online text-to-image tools have been used, and this selection aims to vary the tools (e.g. LookX AI focuses on architectural content), thus varying the generated images, providing the possibility for the study participants to try these tools by choosing free available ones. As well as due to their friendly user interface. The tools are; DreamStudio AI (*DreamStudio*, n.d.), Microsoft Bing AI (*Bing*, n.d.), Limewire AI specifically the BlueWillow model (*LimeWire AI Studio*, n.d.), LookX AI (*LookX AI Cloud*, n.d.) and Leonardo AI (*Leonardo.Ai*, n.d.). The models underlying these tools differ. For instance, DreamStudio AI uses Stable diffusion model (Rombach et al., 2022b) and Microsoft Bing AI is based on DALLE-3 (Betker et al., 2023).

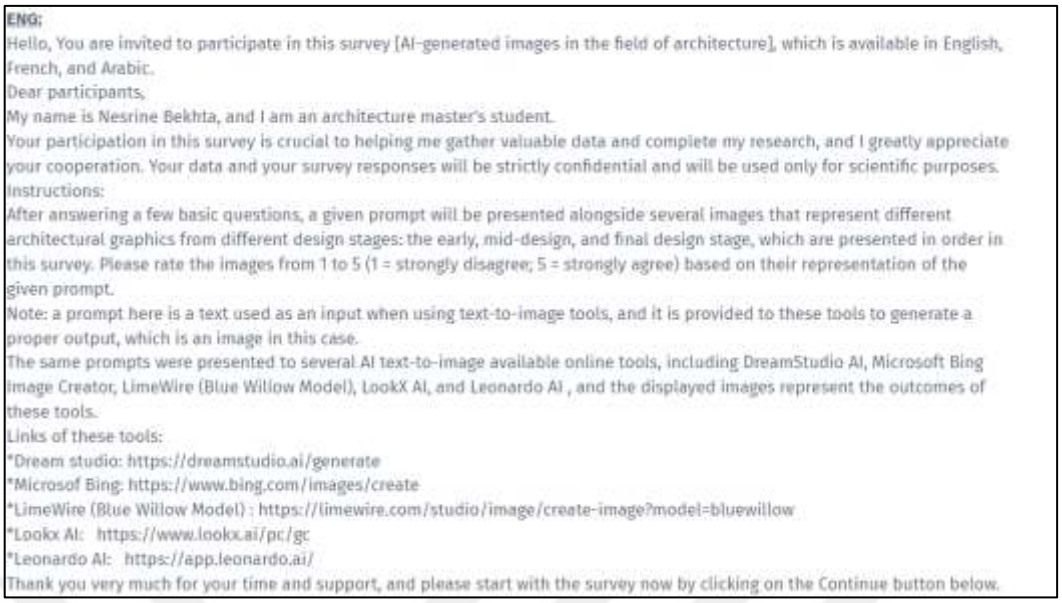
To generate images, the default parameters for each tool have been kept, thus avoiding any confusion, and since they are not modifiable in all tools, to assess fairness among the tools and provide architects without previous knowledge of these technologies the chance to use them in their workflow. For each tool and each architectural representational technique, the same prompt was used. While Bing AI, chosen due to its use of one of the last text-to-image models and its available documentation, generates mandatory four images per prompt, other tools provide the choice of the number of generated images, to ensure fairness, two images are generated from each tool, at once, and for Bing AI two generated images have been selected. Choosing to select two generated images among other possibilities is to balance between exploring the tools' potential, thus providing participants with a sufficient variety as well as minimizing the bias during the rating process in the survey. Thus, for each tool and each representational technique, we have two images, and since there are five tools, it makes the total number for each technique ten images, a sample of the generated images is shown in (Figure 3.3) and all the generated ones are shown from (Figure A1) to (Figure A8).

Figure 3.3: A Sample of the Images Generated by the Different text-to-image Tools Representing an Architectural Bubble Diagram (Created by the Author).

3.3 DATA COLLECTION: A SURVEY

To carry out the study research, a survey was conducted following the generation of the desired images for each architectural representational technique, in order to evaluate the generated images of text-to-image tools which aligns with the approach of the survey conducting as in the case of similar some previous studies such as the work of Saharia et al. (2022) and the work of Marcus et al. (2022).

The survey is structured into various parts. The introduction section holds key information about the study for the purpose of guiding the participants, by containing the importance of participation in this research and the confidentiality of their data. This introduction part also includes the task of this survey, which is to rate the different images generated by different text-to-image AI, using a Likert scale from 1 to 5, where 1 indicates strongly disagree and 5 indicates strongly agree, with the possibility of assigning the same rating for various images. The definition of 'text prompt' is mentioned as well to achieve clarity on the core subject and links to each tool are provided, thus, participants get the chance to try these tools, specifically those who have never experienced them. A sample of the introduction part is shown in (Figure 3.4).



ENG:
Hello, You are invited to participate in this survey [AI-generated images in the field of architecture], which is available in English, French, and Arabic.
Dear participants,
My name is Nesrine Bekhta, and I am an architecture master's student.
Your participation in this survey is crucial to helping me gather valuable data and complete my research, and I greatly appreciate your cooperation. Your data and your survey responses will be strictly confidential and will be used only for scientific purposes.
Instructions:
After answering a few basic questions, a given prompt will be presented alongside several images that represent different architectural graphics from different design stages: the early, mid-design, and final design stage, which are presented in order in this survey. Please rate the images from 1 to 5 (1 = strongly disagree; 5 = strongly agree) based on their representation of the given prompt.
Note: a prompt here is a text used as an input when using text-to-image tools, and it is provided to these tools to generate a proper output, which is an image in this case.
The same prompts were presented to several AI text-to-image available online tools, including DreamStudio AI, Microsoft Bing Image Creator, LimeWire (Blue Willow Model), LookX AI, and Leonardo AI , and the displayed images represent the outcomes of these tools.
Links of these tools:
*Dream studio: https://dreamstudio.ai/generate
*Microsof Bing: https://www.bing.com/images/create
*LimeWire (Blue Willow Model) : https://limewire.com/studio/image/create-image?model=bluewillow
*Lookx AI: https://www.lookx.ai/pc/gc
*Leonardo AI: https://app.leonardo.ai/
Thank you very much for your time and support, and please start with the survey now by clicking on the Continue button below.

Figure 3.4: A Sample of the English Introduction Part of the Survey (Created by the Author).

Participating in the survey is restricted to individuals with an architectural background, thus ensuring their knowledge about the field of architecture and the various representational techniques. The survey is available in English, Arabic, and French to maximize participation. Thus, after the consent text, participants start the answering phase.

After answering basic questions related to their background and their position in the field of architecture (postgraduate architecture students, or architects), participants are asked to answer if they have already used text-to-image tools in the field of architecture and then describe their experience with the text-to-image AI tools in the field ranging from very dissatisfied and very satisfied in order to assess how these tools are being adapted in the field. Raters start the main part of the survey which is the rating process of the generated images, noting that they are not presented in order and the name of the used tools is not shown to avoid any possible bias, each group of images belonging to a specific technique is anteceded by a definition of that technique, in addition to the text prompt used to generate the images. The presented questions that contribute to the research questions are shown in (Table 3.1), and a sample of the survey can be seen in (Figure 3.5).

Table 3.1: The Survey Questions (Created by the Author).

Questions	Question text	Images to rate	Answers
The use of text-to-image tools in the field of architecture	Have you ever used text-to-image tools in the field of architecture?	None	Yes ☐ No☐
	If the answer is yes, how do you describe your experience?		Very dissatisfied/ Dissatisfied Neutral/ Satisfied/ very Satisfied
Images rating for each technique	Rate the following images according to their representation of the following prompt		Rating scale: 1 2 3 4 5
Bubble diagram	Architectural bubble diagram of a two-floor house project		
Conceptual sketches	Conceptual sketches of a two-floor house project		

Table 3.1: The Survey Questions “Table Continued”.

Floor plan drawings	Floor plan drawings of a two-floor house project		
Cross-section drawing	Cross section drawing of a two-floor house project		
Elevation drawing	Elevation drawing of a two-floor house project		
exploded view diagram	Architectural exploded view diagram of a two-floor house project		
Rendered 3D model	Rendered 3D model of a two-floor house project		
Details drawings	Technical detailed staircase drawings of a two-floor house project		

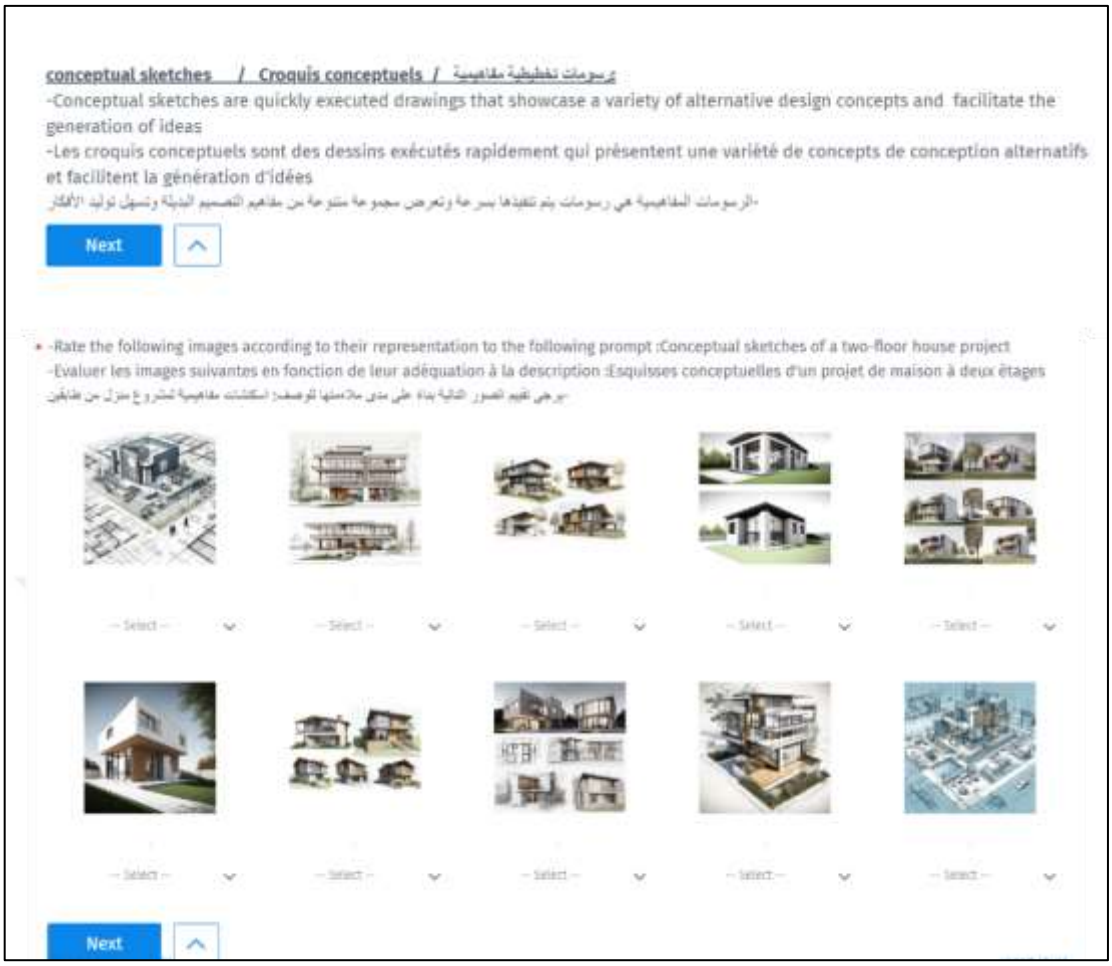


Figure 3.5: A Sample of the Survey for the Conceptual Sketches (Created by the Author)..

3.3.1 Data Analysis

The collected data for this research consists of the first phase that is based on the responses of participants about their experience with the different text-to-image tools. Thus, exploring the widespread use of these tools and how architects are experiencing them.

In the second phase, the rating average scores collected through the survey are analyzed. To find the average score recorded for each technique, the average scores for the images belonging to a specific technique are calculated and they are shown in (Table A1). Then for each technique, the average scores for the two images generated by each tool are counted to explore the score recorded by each tool to assess how these different tools process the various techniques and explore their performance (Table A2). In the aim of investigating the consistency of the same tool for the same text prompt, the scores of the two images generated

by the same tool are compared as seen form thus comparing each average score recorded for each image within the same technique (Table A3) to (Table A11).

4. FINDINGS

The numerical data collected from this survey will shape the findings of this study aiming to gather valuable insights about the current state of using text-to-image tools in terms of generating different architectural representational techniques. The findings of this study are based on numerical scores that are rounded to two decimal places in an attempt to present the findings clearly.

4.1 TEXT-TO-IMAGE TOOLS IN ARCHITECTURAL PRACTICE

At the beginning of the survey, in order to investigate how text-to-image tools are currently interacting with the field of architecture, architects were asked if they already use text-to-image tools in architecture, and how they describe their experience. Among 111 architects who arrived at this question, 39 architects (35.14%) answered 'yes' and 72 architects (64.86%) answered 'No'.For the ones who answered 'yes' only 27 architects expressed their experience, when a difference is seen in how architects experienced these tools, most of them (40.74%) describe it as 'neutral', (37.04%) as 'satisfied', (11.11%) as 'very satisfied', (11.11%) as 'dissatisfied' and no one described it as 'very dissatisfied', **Figure 4.1** shows the experiences of architects gathered using a Likert scale.

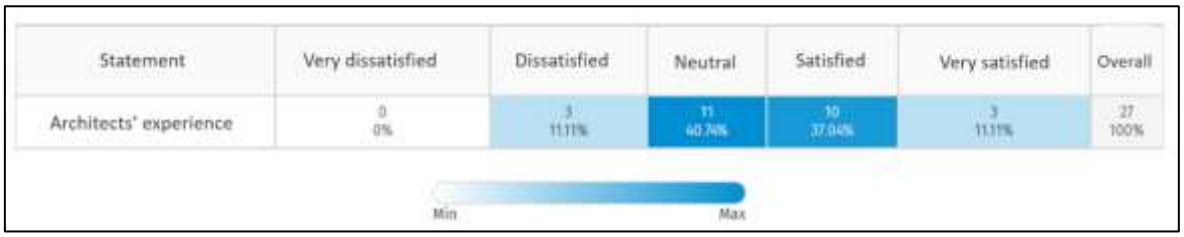


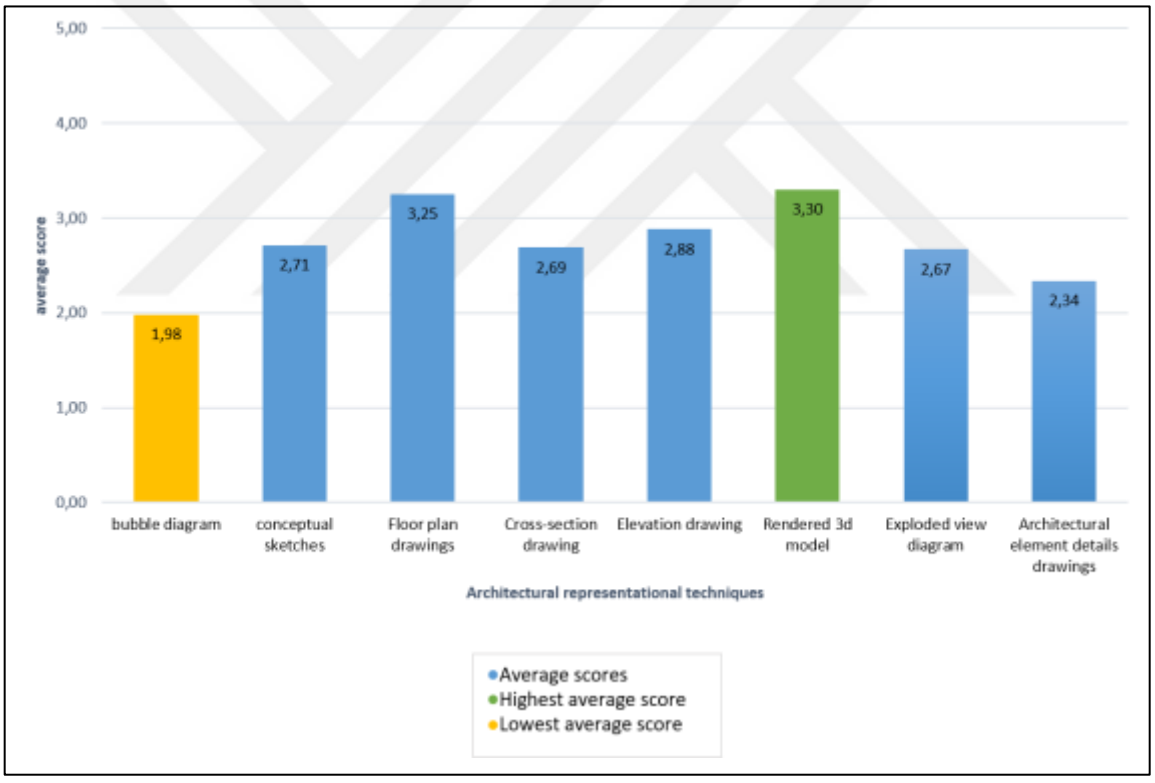
Figure 4.1: The Experience of Architects With Text-to-image Tools In the Field of Architecture (Created by the Author).

4.2 THE PERFORMANCE OF TOOLS ACROSS THE DIFFERENT TECHNIQUES

To assess the performance of the different text-to-image tools in generating the different architectural representational techniques, the average score of each technique has been revealed, and then compared with the other scores. The results show that there is a distinct preference for the participants for some images representing the different representational

techniques as seen in (Chart 4.1) when the rendered 3D model scored highest with an average score of 3.30 (out of 5) suggesting the visual appeal of the technique, as well as its easiness to be conveyed using text input, conversely, the lowest average score, was scored for the images representing the bubble diagram techniques with an average score of 1.98 out of 5 suggesting the complexity of text-to-image tools to convey the idea of bubble diagram using text description. The average scores recorded for the other techniques have been revealed as well, and in descending order, they are floor plan drawings, elevation drawing, conceptual sketches, cross-section drawing, exploded view diagram, and architectural element details drawings.

Chart 4.1: The Average Score Recorded for the Different Architectural Representational Techniques (Created by the Author).

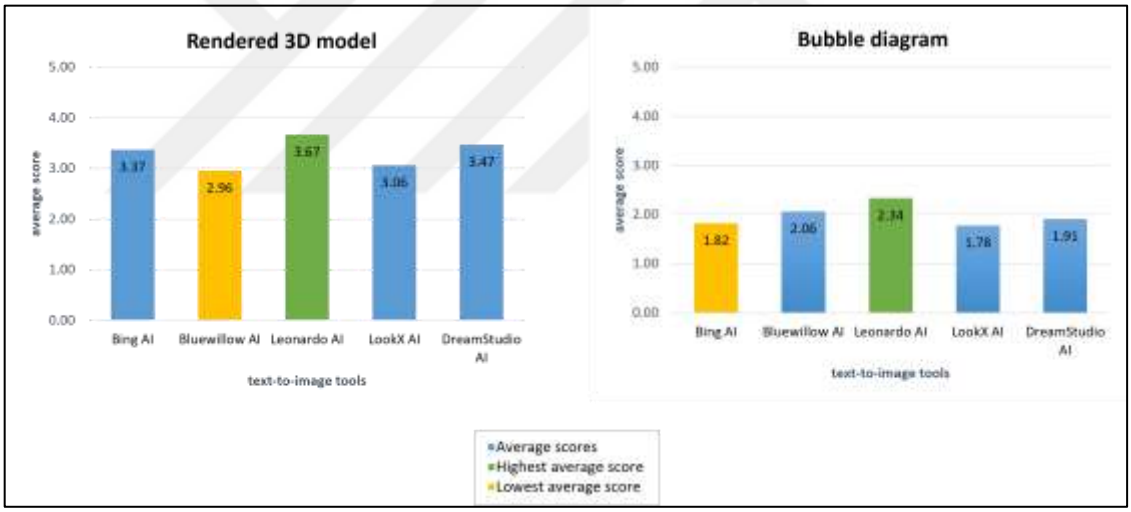


4.3 THE PERFORMANCE OF TOOLS FOR EACH TECHNIQUE

The findings of the performance of tools across the different techniques lead to the exploration of the average scores recorded for each tool and the way they interact with the different representational techniques to see how they behave to the same text input for every technique. The results show that each tool operates differently for the various architectural

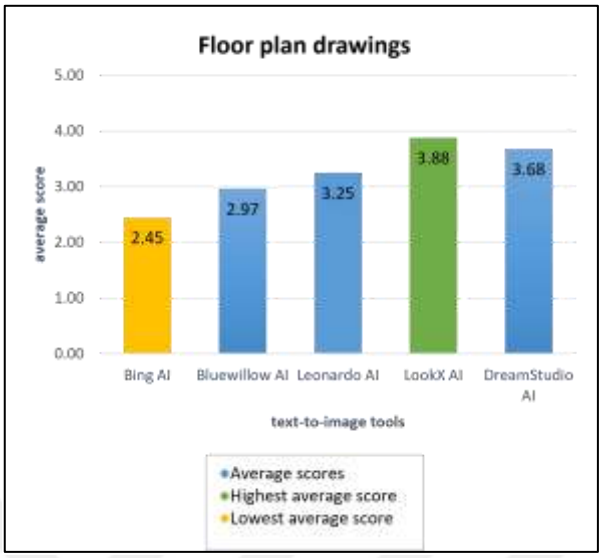
representational techniques with variability in performance; there is no single tool that performs as the most efficient one in handling all the different representational techniques. For instance, for the highest score recorded by the rendered 3D model, all text-to-image tools scored higher than 3 except for one tool that scored 2.96 (out of 5), which is the lowest score recorded by BluWillow AI, and the Leonardo AI tool scored highest with a score of 3.67. For the lowest score recorded by the bubble diagram, all the scores are under 2.5, and Leonardo AI tool scored highest too with a score of 2.34 and Bing AI scored the lowest as (Chart 4.2) shows, even though the average scores for each technique are different, Leonardo AI scored the highest for both techniques.

Chart 4.2: Chart of The Average Scores Recorded for Each Tool for the Rendered 3D Model and the Bubble Diagram. Left: The Recorded Scores for the Rendered 3D-Model Representational Technique. Right: The Recorded Scores for the Bubble Diagram Representational Technique (Created by the Author).



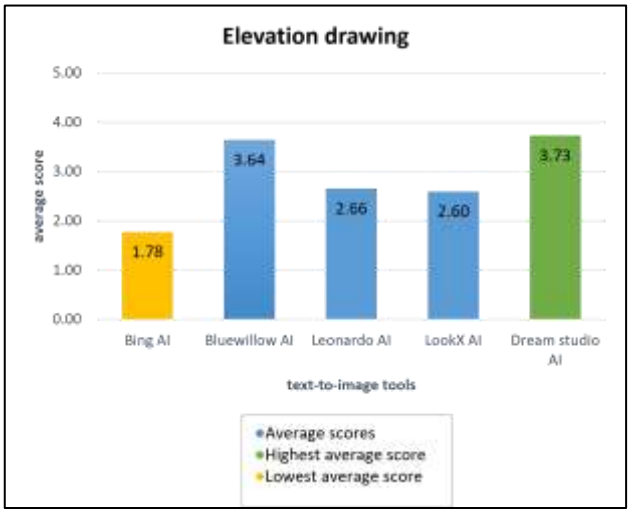
For the floor plan drawings scored 3.25, it can be seen from (Chart 4.3) that there is a slight discrepancy in the average scores of the tools, with the highest score recorded for the LookX AI tool and the lowest score has been recorded for the Bing AI tool.

Chart 4.3: The Average Scores for Each Text-to-image Tool Recorded for the Floor Plan Drawings (Created by the Author).



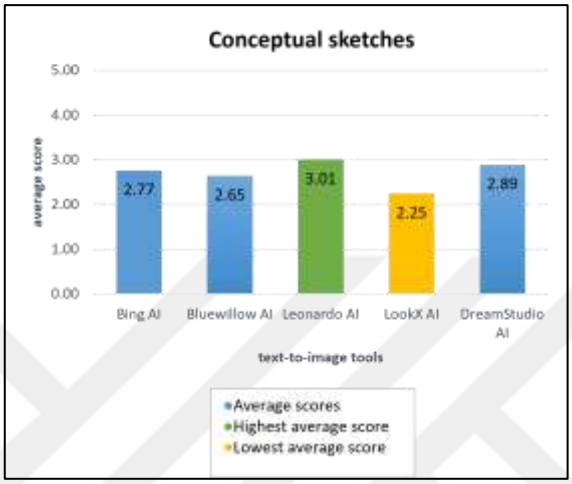
The elevation drawing scored 2.88, and it can be seen from (Chart 4.4) that there is a fairly large discrepancy in the score recorded for each tool, noting that DreamStudio AI scored the highest. Bing AI tool scored the lowest, as in the case of the bubble diagram and the floor plan drawings.

Chart 4.4: The Average Scores For Each Text-to-image Tool Recorded for the Elevation Drawing (Created by the Author).



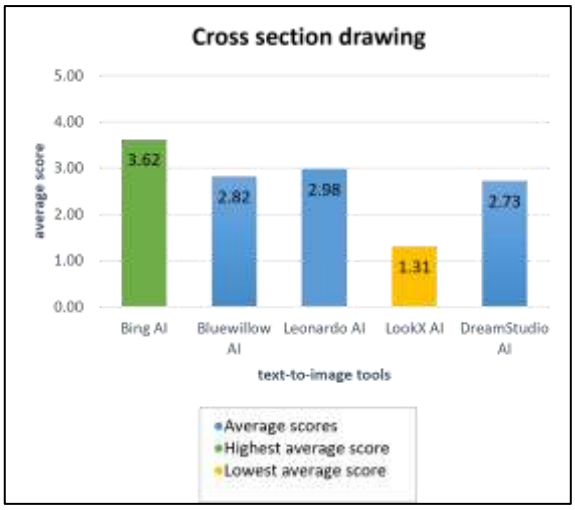
The conceptual sketches recorded 2.71, mentioning the slight discrepancy in the average scores of the tools seen in (Chart 4.5), and Leonardo AI tool scored the highest, while LookX AI scored the lowest, although it scored the highest for the floor plan drawings.

Chart 4.5: The Average Scores for Each Text-to-image Tool Recorded for the Conceptual Sketches (Created by the Author).



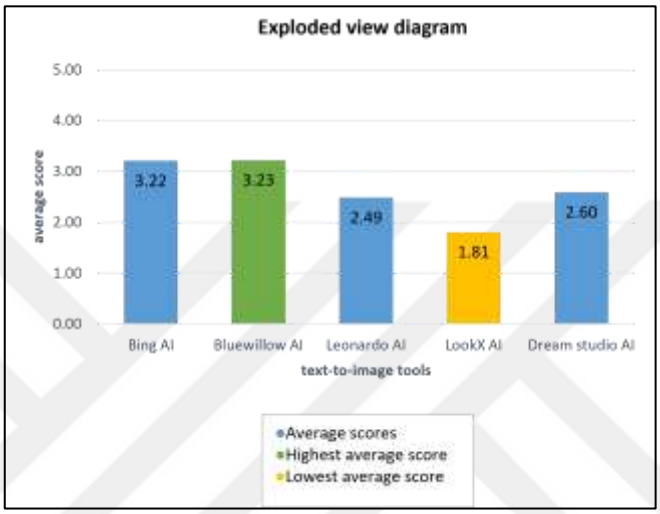
For the cross-section drawing, it can be seen from (Chart 4.6) that there is a discrepancy between the scores recorded for each tool specifically between Bing AI tool and LookX AI tool. Bing AI tool scored the highest even though it scored the lowest for some other techniques such as the elevation drawing, floor plan drawings, and the bubble diagram. LookX AI scored the lowest even though it scored the highest for the floor plan drawings.

Chart 4.6: The Average Scores for each Text-to-image Tool Recorded for the Cross-Section Drawing (Created by the Author).



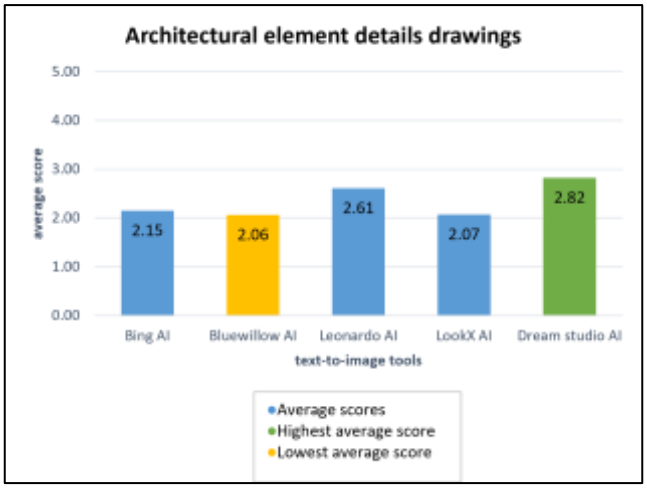
The exploded view diagram shows a difference in average scores recorded for each tool as (Chart 4.7) shows, noting that BlueWillow AI scored the highest although it scored the lowest for the rendered 3D model and the element details drawings, and LookX AI scored the lowest again, as in the case for some other techniques.

Chart 4.7: The Average Scores for Each Text-to-image Tool Recorded for the Exploded View Diagram (Created by the Author).



For the architectural element details drawing, the difference between the scores recorded by each tool is a bit slight as (Chart 4.8) shows, with the tool DreamStudio AI scoring the highest as it is the case for the elevation drawing, and BlueWillow scored the lowest as it is the case for the rendered 3D model while it scored highest for the exploded view diagram.

Chart 4.8: The Average Scores for Each Text-to-image Tool Recorded for the Architectural Element Details (Created by the Author).

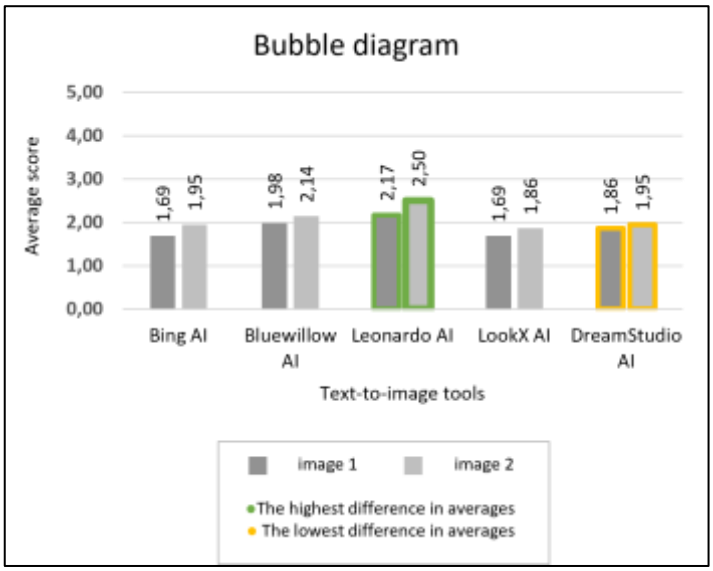


4.4 THE RESPONSE OF THE TOOLS TO THE SAME TEXT PROMPT

The reaction of the same tool to the same text prompt has been investigated to explore the consistency of the same tool for the same text prompt by exploring and comparing the average scores of the two images generated from each tool at once for the same text prompt within the representational techniques, meaning examining the difference between the average scores of the images that are seen as highly or slightly noticeable and could differ from a tool to another or even the same tool across the various representational techniques.

For the images representing the bubble diagram and generated by the different tools it can be seen from (Chart 4.9) that there is a slight difference between the average scores of each two images generated by one tool when the highest difference in average scores is found between the images generated by Leonardo AI tool when one image scored 2.17 (out of 5) and the other one scored 2.50 with a difference of 0.33 (6.6٪) and the lowest difference is found between the scored of the images generated using DreamStudio AI tool when one image scored 1.86 and the other one scored 1.95 with a difference of 0.09 (1.8٪).

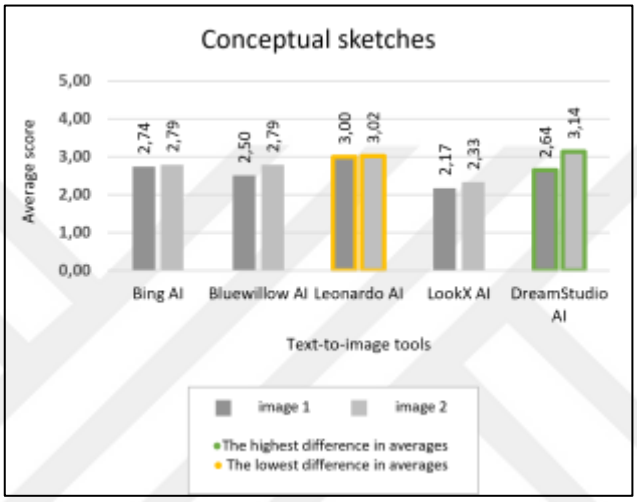
Chart 4.9: The Average Scores for Every Two Images Generated by Each of the Text-To-Image Tools for the Bubble Diagram (Created by the Author).



For the conceptual sketches, from (Chart 4.10), it is observed that there is a slight difference between the average scores of each of the two images generated by one tool, noting that in this case unlike the case of the bubble diagram, the highest difference is recorded between

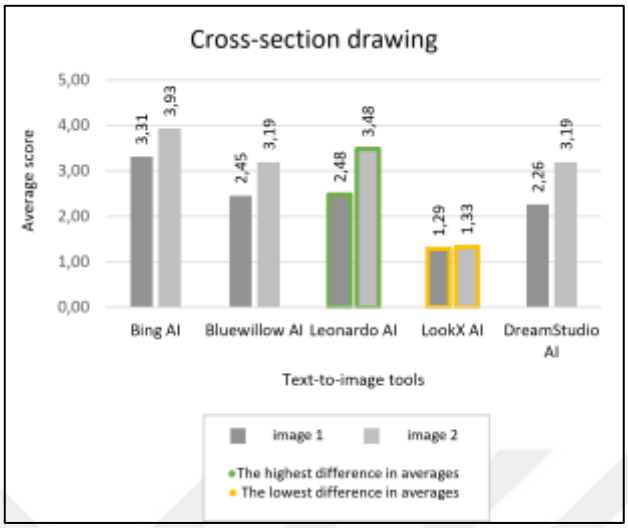
the average scores recorded for images generated by DreamStudio AI tool when one image scored 2.64 and the other one scored 3.14 with a difference of 0.5 (10٪) and the lowest difference is recorded between the average scores recorded for images generated by Leonardo AI tool when one image scored 3.00 and the other one scored 3.02 with a difference of 0.02 (0.4٪).

Chart 4.10: The Average Scores for Every Two Images Generated by Each Text-to-image Tool for the Conceptual Sketches (Created by the Author).



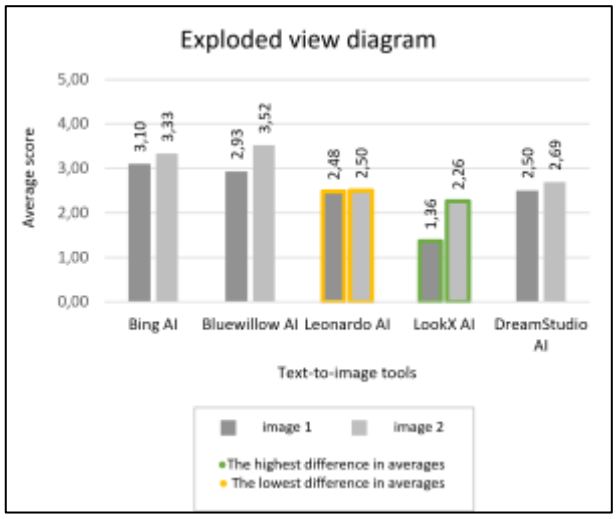
In the case of cross-section drawing, (Chart 4.11) shows that the difference in the average scores between the two images generated by each tool is significant, thus, the consistency is low somewhat except for the difference between the average scores of the images generated by LookX AI tool which is the lowest, when one image scored 1.29 and the other one scored 1.33 with a difference of 0.04 (0.8٪), and the biggest difference is found between the images generated using Leonardo AI tool when one image socred 2.48 and the other one scored 3.48 with the difference of 1 (20٪).

Chart 4.11: The Average Scores for Every Two Images Generated by Each Text-To-Image Tool for the Cross-Section Drawing (Created by the Author).



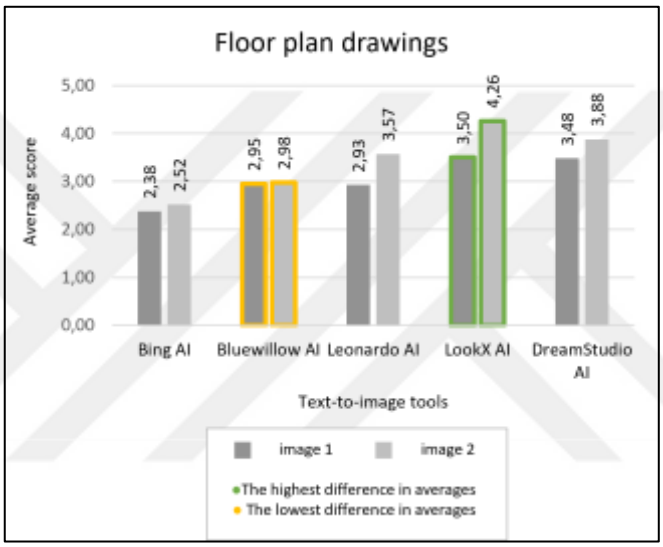
For the exploded view diagram, as seen in (Chart 4.12), the difference in the average scores between every two images generated by the same tool is a bit unstable for some tools, the difference in scores for the images generated by the LookX AI tool is the highest, one image scored 1.36 and the other one scored 2.26, meaning that the difference is 0.9 (18٪), and the difference in scores for the images generated by Leonardo AI is the lowest when one image scored 2.48 and the other one scored 2.50 with a difference of 0.02 (0.4٪), which is the opposite of the cross-section drawing case.

Chart 4.12: The Average Scores for Every Two Images Generated By Each Text-to-image Tool for the Exploded View Diagram (Created by the Author).



In the case of the floor plan drawings, it is observed from (Chart 4.13) that the difference between the average scores between every two images generated by the same tool is a bit unstable for some tools, the highest difference in scores is seen in the two images generated by LookX AI tool when one image scored 3.50 and the other scored 4.36 with a difference of 0.76 (15.2٪) and the lowest difference is found between the two images generated by BlueWillow AI tool with a difference of 0.03 (0.6٪).

Chart 4.13: The Average Scores for Every Two Images Generated by Each Text-to-image Tool for he Floor Plan Drawings (Created by the Author).



In the case of the elevation drawing, it can be seen from (Chart 4.14) that the difference in the average scores between every two images generated by the same tool, is unstable when the difference in scores between the images generated by Leonardo AI and LookX AI is very significant. The difference in scores of the images generated by LookX AI is the highest among others, one image scored 1.69 and the other one scored 3.50 with a difference of 1.81(36.2٪), and the difference in scores using Bing AI is the lowest, one image socred 1.69 and the other scored 1.86 with a difference of 0.17 (3.4٪). The elevation drawing recorded the biggest difference between the highest and lowest difference between each two images, and **(**Figure 4.2) shows these images generated by Bing AI and LookX AI.

Chart 4.14: The Average Scores for Every Two Images Generated by Each Text-to-image Tool for the Elevation Drawing (Created by the Author).

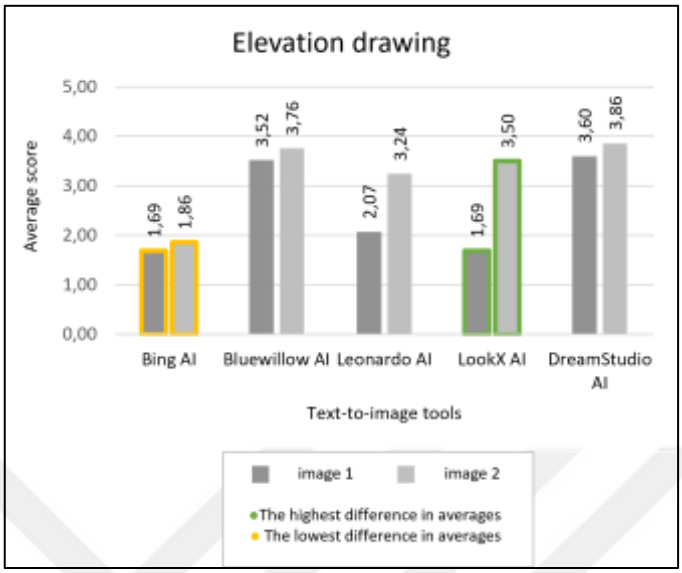
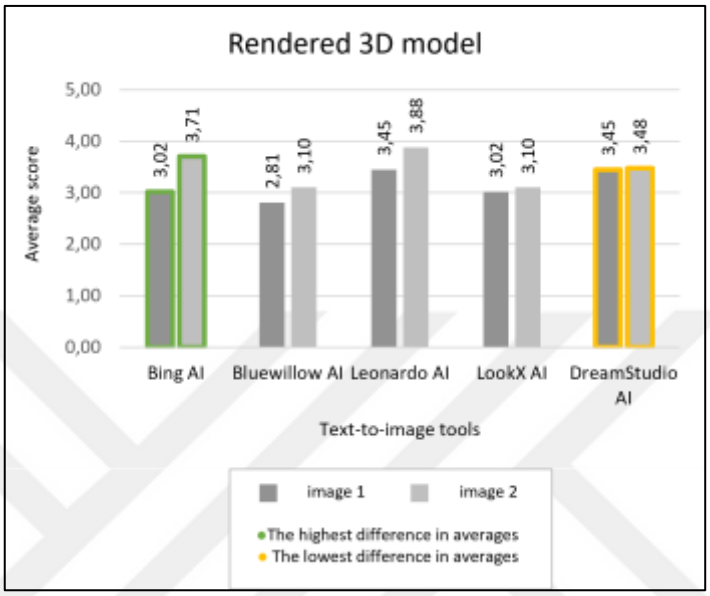


Figure 4.2: The Images Generated by LookX AI and Bing AI Representing the Elevation Drawing, and Their Average Score (Created by the Author).

For the rendered 3D model, (Chart 4.15) shows that the difference in the average scores between every two images generated by the same tool is slight somewhat and the highest one is found between the images generated using Bing AI tool in this case when one image scored 3.02 and the other one scored 3.71 with a difference of 0.69 (13.8٪) and the lowest

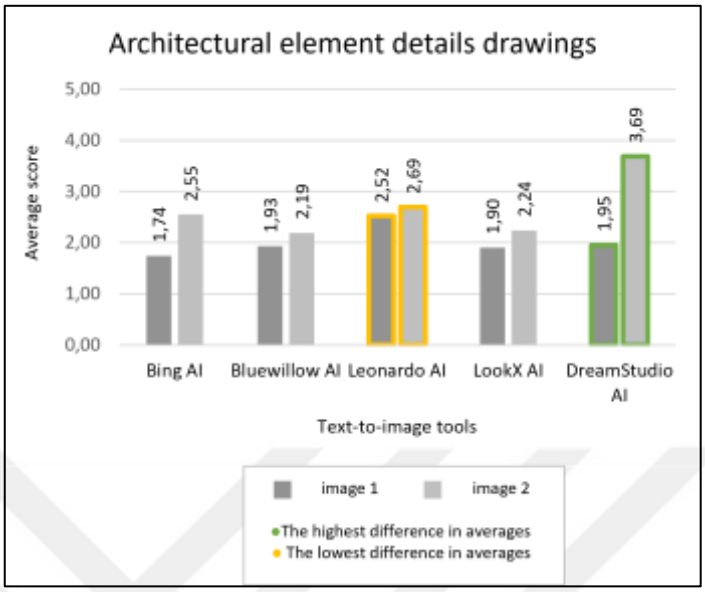
score is found between the images generated by DreamStudio AI, one image scored 3.45 and the other one scored 3.48 with a difference of 0.03 (0.6٪).

Chart 4.15: The Average Scores for Every Two Images Generated By Each Text-To-Image Tool for the Rendered 3D Model (Created by the Author).



In the case of the architectural element details drawings, (Chart 4.16) shows that the difference in the average scores between every two images generated by the same tool is a bit unstable for some tools, specifically between the images generated by DreamStudio AI which is the highest difference when one image scored 1.95 and the second one scored 3.69 with a difference of 1.74 (34.8٪), and the lowest difference has been found between the images generated by Loenardo AI, one image scored 2.52 and the second one scored 2.69 with a difference of 0.17 (3.4٪).

Chart 4.16: The Average Scores for Every Two Images Generated by Each Text-To-Image Tool for the Architectural Element Details Drawings (Created by the Author).



The findings collected by analyzing the data gathered through the study survey are presented in (Table 4.1).

Table 4.1: The Study Findings at a General and a Specific Level (Created by the Author).

The use of text-to-image tools	Analyzed data	Results
General use of text-to-image tools in the architectural workflow	The utilization of text-to-image tools	Text-to-image tools are being adopted significantly by architects
	The experience of architects with text-to-image tools	Participant architects appear to have a good to neutral experience with a small group voting for a dissatisfied and very satisfied experience

Table 4.1: The Study Findings at a General and a Specific Level “Table Continued”

Specific use of text-to-image tools in generating architectural representational techniques	The performance of the tools across the techniques	-Variability in terms of generating the techniques - A preference of participants for images representing some techniques over others
	The performance of tools for each technique	-Each tool operates differently for the various architectural representational techniques -No single tool performs the most efficiently in handling all the different representational techniques
	The consistency of the tools	-The consistency of the tools differs and it is either slightly or highly noticeable from one technique to another / the same tool across the different techniques

5. DISCUSSION

The findings of this study can aid architects gain valuable insights into the usefulness of using text-to-image in the field of architecture, specifically in generating images representing different architectural representational techniques, in terms of the difference in tool's performance, the processing of the various techniques using the different tools and their consistency.

The results show that a significant number of architects who answered about whether they already used text-to-image tools in the field of architecture, answered with 'yes' indicating that these technologies are being adopted in the field of architecture significantly. The results of how architects describe their experience with these tools indicate that architects appear to have a good experience to neutral somewhat with a small group of architects who describe it as 'very satisfied' and 'dissatisfied', this latter may indicate the misunderstanding of the textual prompt that text-to-image tools may face, (Figure 5.1) shows some 'misunderstanding' cases in interpreting the textual prompt for this thesis study case at the same time it shows some samples of some correct outputs. Worth noting that none of the architects describe their experience as 'very dissatisfied' demonstrating that these tools are not undesired. However, human intervention is still important which aligns with (Radhakrishnan, 2023), (Z. Zhang et al., 2023), and (Tanugraha, 2023).

Figure 5.1: Samples of the Textual Prompt Output. Top: 'Misunderstanding' Cases in Interpreting the Textual Prompt. Bottom: Samples of the Correct Output Using the Same Prompt (Created by the Author).

The results of the text-to-image tools' performance indicate the variability of this latter in terms of generating images representing the different architectural representational techniques, suggesting that their performance depends on the nature of the technique, how architects used to generate it, as well as the performance could be affected by the data conveyed using the text descriptions. The highest average score recorded for the images representing the rendered 3D model could indicate the easiness of the text-to-image tool to deal with this representational technique, as it may indicate the familiarity of architect in using the numerical tools when dealing with this kind of representational technique, and it suggests also that the idea of the rendered 3d model can be deciphered with ease through text description, whereas, the lowest average score recorded for the images representing an architectural bubble diagram can be interpreted as a result of the difficulty for the text-to-image tool to handle it due to its nature that could be more abstract than the rendered 3D model and usually architects deal with it using conventional tools to express their ideas, also,

the text description may affect when conveying the bubble diagram technique and it may necessitate more data covering more details about the project, its environment, and more information that may overpass a sole text input.

The performance variance factors also might be applied to the other techniques. For instance, the floor plan drawings recorded a high score may suggest that the images could convey the overall representation of the plans, even though, not all the generated plans convey the semantic features or take into consideration the whole description in the prompt such as the plural term ‘drawings’, when some images represent only one plan drawing. which is the same case for the conceptual sketches. For the elevation drawing, there is dissimilarity in the average scores of the images generated by each tool due to the differences between them in representing ‘ a two-floor house project’ and understanding the prompt properly. Cross-section drawing recorded the lowest score among the orthographic projection, due to its nature of depending on other visual elements to effectively interpret the project’s information as well as its technical and graphical features that may not be completely conveyed through text descriptions, especially for a specific case study, likewise in the case of the exploded view diagram and the architectural element details drawings when a low score has been recorded with a significant variation in the tools’ average scores. The results shed light on how the performance of text-to-image tools may vary in generating the different representational techniques when each technique may be conveyed through text differently, based on its nature, required details, and the way architects used to generate it.

The results of comparing the average scores of the different images generated by the various tools to investigate their reaction to the same text prompt, revealed that every single tool behaves and handles the representational techniques in an uneven manner when one tool can perform better in terms of generating images representing a specific representational technique and struggles to perform the same way for another technique, this might be explained by the variety found in the models behind these tools as they form the backbone of text-to-image tools, as well as the data used to train the tools may differ and a tool can be trained on images representing a technique more than another, which suggests using different tools to find the most accurate one in generating images that represent the desired representational techniques and their specific aspects in addition to fine tuning the model

with a specific dataset of the desired representational technique which is possible for some tools.

The response of the same tool to the same text prompt has been examined to explore the consistency of the same tool for the same text prompt. Appealingly**,** the results indicate that even though the images are generated at once using the same text prompt and for the same representational techniques, the consistency of the tools differs, which might be explained by different factors that are the difference in tools' parameters which are not modifiable in all tools (e.g. Bing AI), instanced by the guidance scale and the seeds and are set in default for this study. The difference may also arise from the perception of the participants which may vary according to their knowledge of the field, but still, the difference in average score can be helpful in the design workflow when the images may present different elements that can be used as an inspirational tool for architects and guide them to make entire new decisions, noting that not all the generated images may represent the prompt entirely, Thus, a careful selection among the generated images based on architects' knowledge is required.

5.1 LIMITATIONS

Despite the steps that have been made to reduce limitations, such as filtring the participants' backgrounds to only the ones who at least had their graduate studies in the field of architecture and providing the survey with three different languages to maximize participation and avoid any misunderstanding with the textual information, the survey ratings represent the perception of the participants on the subject which may differ from an architect to another and may be considered subjective and sensitive to biases, as well as the difference of the names of representational techniques that may vary. For instance, an elevation can also be called a 'façade'.

The statistical results derived from the survey may be affected by the size of the sample when the number of participants who completed the rating process is forty-two participants among two hundred fifty total responses. Moreover, the details of the survey may not be fully taken into consideration by the participants, who may neglect to capture some aspects of the survey's details such as the prompts' details within the text description of the floor plan drawings when the word 'drawing' is used in the plural 'drawings' intentionally, as well as the aspects ad background of the images may affect their rating.

The study includes a sample of five text-to-image online available tools, which represent a small portion of the available ones in a specific time era that may behave differently. Furthermore, it is vital to acknowledge that the field is a growing one, and many tools with different capabilities may have appeared. To ensure fairness among the tools, the default parameters as well as the specific model within the tool and the number of images per tool were set in default in this study, during the research period, which may present a limitation.

Text-to-image tools have different underlying architectures, that are not publicly available for all the tools. For instance, I tried to contact the user service for the tools that do not have scientific articles or do not provide information about their models, such as Bluewillow AI, and LookX AI, when I got a general answer and I did not get anyone for the Leonardo AI tool, acknowledging that knowing their models and their training data may help to get more extract more valuable findings and emphasize the other ones.

The study focuses on a group of representational techniques using different text-to-image tools within the case study of a two-floor house project, meaning that the results derived from the survey may differ for other types of architectural projects, and may not be standardized for the whole architectural workflow. Furthermore, the study relies on text prompts, and the evaluations are based on images generated using a single text prompt without further iterations, which may limit the utilization of full advantage of the tools.

6. CONCLUSION

In this thesis, the topic of using text-to-image AI tools in the field of architecture, specifically in generating various architectural representational techniques has been covered, to illuminate the perspectives of using these state-of-the-art tools in the architectural workflow by revealing their role and their applicability.

Based on a systematic approach, this study offers important observations. The literature review section provides a hierarchical comprehension of the topic, starting with an overview view, after which the details are presented, by explaining different terms related to text-to-image generative tools, we highlight Artificial intelligence, its related terms, and subfields, generative AI, and text-to-image working modes, thus, exploring broad literature which contributes the understanding of the topic from a broader scope.

The methodology for this study is based on data creation in the first phase, which is represented by generating images using different text-to-image tools that are available online and for various architectural representational techniques which are bubble diagram, conceptual sketches, floor plan drawings, cross-section drawing, elevation drawing, exploded view diagram, rendered 3D model and architectural element details with the aid of created standardized prompts to ensure fairness among the techniques, and then these images are shared on a survey targeting architects to rate them after answering basic questions related to their position in architecture and about their experience with text-to-image tools in the realm of architecture.

The findings illustrate different aspects of text-based image-generation tools. On a general level and according to participants, these tools are being adopted in the field of architecture and the experience of participant architects appears to be neutral to positive experience which shows that these tools are being adopted significantly in the architectural and some dissatisfied experiences may arise from the fact that these tools may misunderstand the prompt entirely and generate images that do not represent the desired outcomes.

On a specific level, generating architectural representational techniques using these tools, reveals that their performance varies, and they handle the different techniques in an uneven manner depending on the aspects of each technique, how it can be conveyed through textual descriptions, and the way architects used to handle each technique in addition to the

underlying models that perform behind each tool and the data that are trained on. The tools also react to the same prompt differently for the same technique when even the images are generated at once using the same textual prompt, their consistency differs and can be described as obvious or slightly remarkable, which suggests a multi-tool approach application to guarantees the desired results while taking into consideration the variety of generated images for the same prompt and how it can or can not convey the idea behind the technique entirely which requires the intervention of human architects that is still crucial.

It is crucial to acknowledge limitations, even though the delivered perspectives from this study. The sample of participants may cover a small portion of perceptions, in addition to the findings that are limited to the selected techniques for this study and may differ for other types of projects with more detailed textual prompts.

To better understand the usefulness of text-to-image tools future research should focus on studying how each technique may be conveyed better through textual description and designing textual templates and prompt generator tools specified for representational technique. Moreover, future research should focus on developing text-to-image tools that take the architectural techniques as a whole by generating a technique based on another one, in addition to the importance of collaboration between text-to-image tools developers and architects that may suggest creating this kind of tools with an interface that categorizes the different architectural representational techniques, would enhance more control.

This study shows the potential of using text-to-image tools and how architects can benefit from them as an inspirational tool during the design workflow while highlighting that human contribution is still important in controlling the output images and these AI tools, up to now can be considered as an ‘assistant’ for architects, the final decision-makers.

REFERENCES

Alammar, J. (2022, October 4). *The Illustrated Stable Diffusion*. http://jalammar.github.io/illustrated-stable-diffusion/

Albaghajati, Z. M., Bettaieb, D. M., & Malek, R. B. (2023). Exploring text-to-image application in architectural design: Insights and implications. *Architecture, Structures and Construction*, *3*(4), 475–497. https://doi.org/10.1007/s44150-023-00103-x

Alberto, D. (1982). *Architectural representation ; spatial comprehension and assessment through visualization technique* [Thesis, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/62892

Allen, S. (2009). *Practice: Architecture, Technique and Representation*. Routledge.

Bandi, A., Adapa, P. V. S. R., & Kuchi, Y. E. V. P. K. (2023). The Power of Generative AI: A Review of Requirements, Models, Input–Output Formats, Evaluation Metrics, and Challenges. *Future Internet*, *15*(8), Article 8. https://doi.org/10.3390/fi15080260

Banh, L., & Strobel, G. (2023). Generative artificial intelligence. *Electronic Markets*, *33*(1), 63. https://doi.org/10.1007/s12525-023-00680-1

Baraheem, S. S., Le, T.-N., & Nguyen, T. V. (2023). Image synthesis: A review of methods, datasets, evaluation metrics, and future outlook. *Artificial Intelligence Review*, *56*(10), 10813–10865. https://doi.org/10.1007/s10462-023-10434-2

Baydogan, M., & Karamış, Z. (2023). A Study on the use of diagrams as a form of representation (expression) in architectural education. *JOURNAL OF AWARENESS*, *8*, 415–439. https://doi.org/10.26809/joa.2108

Betker, J., Goh, G., Jing, L., TimBrooks, Wang, J., Li, L., LongOuyang, JuntangZhuang, JoyceLee, YufeiGuo, WesamManassra, PrafullaDhariwal, CaseyChu, YunxinJiao, & Ramesh, A. (2023). *Improving Image Generation with Better Captions*. https://www.semanticscholar.org/paper/Improving-Image-Generation-with-Better-Captions-Betker-Goh/cfee1826dd4743eab44c6e27a0cc5970effa4d80

Bhardwaj, A., Di, W., & Wei, J. (2018). *Deep Learning Essentials: Your hands-on guide to the fundamentals of deep learning and neural network modeling*. Packt Publishing Ltd.

Bianchi, F., Attanasio, G., Pisoni, R., Terragni, S., Sarti, G., & Lakshmi, S. (2021). *Contrastive Language-Image Pre-training for the Italian Language* (arXiv:2108.08688). arXiv. https://doi.org/10.48550/arXiv.2108.08688

Bing. (n.d.). Bing. https://www.bing.com/images/create/

Bodnar, C. (2018). *Text to Image Synthesis Using Generative Adversarial Networks*. https://doi.org/10.13140/RG.2.2.35817.39523

Brisco, R., Hay, L., & Dhami, S. (2023). EXPLORING THE ROLE OF TEXT-TO-IMAGE AI IN CONCEPT GENERATION. *Proceedings of the Design Society*, *3*, 1835–1844. https://doi.org/10.1017/pds.2023.184

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., & Sun, L. (2023). *A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT* (arXiv:2303.04226). arXiv. https://doi.org/10.48550/arXiv.2303.04226

Chaudhary, D., & Vasuja, Er. R. (2019). A Review on Various Algorithms used in Machine Learning. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*. https://www.academia.edu/44802130/A_Review_on_Various_Algorithms_used_in_Machine_Learning

Chen, J., Shao, Z., & Hu, B. (2023). Generating Interior Design from Text: A New Diffusion Model-Based Method for Efficient Creative Design. *Buildings*, *13*(7), Article 7. https://doi.org/10.3390/buildings13071861

Chen, J., Wang, D., Shao, Z., Zhang, X., Ruan, M., Li, H., & Li, J. (2023). Using Artificial Intelligence to Generate Master-Quality Architectural Designs from Text Descriptions. *Buildings*, *13*(9), Article 9. https://doi.org/10.3390/buildings13092285

Croitoru, F.-A., Hondru, V., Ionescu, R. T., & Shah, M. (2023). Diffusion Models in Vision: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(9), 10850–10869. https://doi.org/10.1109/TPAMI.2023.3261988

de la Fuente Suárez, L. (2016). Towards experiential representation in architecture. *Journal of Architecture and Urbanism*, *40*, 47–58. https://doi.org/10.3846/20297955.2016.1163243

Delipetrev, B., Tsinaraki, C., & Kostic, U. (2020). *Historical Evolution of Artificial Intelligence* [Monograph]. Publications Office of the European Union. https://publications.jrc.ec.europa.eu/repository/handle/JRC120469

DreamStudio. (n.d.). https://beta.dreamstudio.ai/generate

Farrelly, L. (2008). *Basics Architecture—Representational Techniques*. Fairchild Books AVA. http://archive.org/details/Basics_Architecture_Representational_Techniques

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. *Business & Information Systems Engineering*, *66*(1), 111–126. https://doi.org/10.1007/s12599-023-00834-7

Fishman, N., Klarner, L., De Bortoli, V., Mathieu, E., & Hutchinson, M. (2024). *Diffusion Models for Constrained Domains* (arXiv:2304.05364). arXiv. https://doi.org/10.48550/arXiv.2304.05364

Forghani, R. (2020). *Machine Learning and Other Artificial Intelligence Applications, An Issue of Neuroimaging Clinics of North America, E-Book: Machine Learning and Other Artificial Intelligence Applications, An Issue of Neuroimaging Clinics of North America, E-Book*. Elsevier Health Sciences.

Gafni, O., Polyak, A., Ashual, O., Sheynin, S., Parikh, D., & Taigman, Y. (2022). *Make-A-Scene: Scene-Based Text-to-Image Generation with Human Priors* (arXiv:2203.13131). arXiv. http://arxiv.org/abs/2203.13131

Ghosh, M., & Thirugnanam, A. (2021). Introduction to Artificial Intelligence. In K. G. Srinivasa, S. G. M., & S. R. M. Sekhar (Eds.), *Artificial Intelligence for Information*

Management: A Healthcare Perspective (pp. 23–44). Springer. https://doi.org/10.1007/978-981-16-0415-7_2

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*. https://proceedings.neurips.cc/paper_files/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139–144. https://doi.org/10.1145/3422622

Graikos, A., Yellapragada, S., & Samaras, D. (2023). *Conditional Generation from Unconditional Diffusion Models using Denoiser Representations* (arXiv:2306.01900). arXiv. https://doi.org/10.48550/arXiv.2306.01900

Gregor, K., Danihelka, I., Graves, A., Rezende, D., & Wierstra, D. (2015). DRAW: A Recurrent Neural Network For Image Generation. *Proceedings of the 32nd International Conference on Machine Learning*, 1462–1471. https://proceedings.mlr.press/v37/gregor15.html

Gui, J., Sun, Z., Wen, Y., Tao, D., & Ye, J. (2023). A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications. *IEEE Transactions on Knowledge and Data Engineering*, *35*(4), 3313–3332. https://doi.org/10.1109/TKDE.2021.3130191

Haenlein, M., & Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*, *61*, 000812561986492. https://doi.org/10.1177/0008125619864925

Hua, T. K. (2022). *A Short Review on Machine Learning*. https://www.authorea.com/users/510271/articles/587406-a-short-review-on-machine-learning?commit=adb3126655cb1e97c77fee92fa68b5461b720f85

Jaruga-Rozdolska, A. (2022). *Artificial intelligence as part of future practices in the architect's work: MidJourney generative tool as part of a process of creating an architectural form*. 95–104. https://doi.org/10.37190/arc220310

Jeffery, O. (2022). *Explain Artificial Intelligence and History of Artificial Intelligence*.

Jiao, L., & Zhao, J. (2019). A Survey on the New Generation of Deep Learning in Image Processing. *IEEE Access*, *7*, 172231–172263. https://doi.org/10.1109/ACCESS.2019.2956508

Jin, M., Zhang, C., Yu, Q., Xue, H., Jin, X., & Yang, X. (2023). *A Simple and Effective Baseline for Attentional Generative Adversarial Networks* (arXiv:2306.14708). arXiv. https://doi.org/10.48550/arXiv.2306.14708

Kelleher, J. D. (2019). *Deep Learning*. MIT Press.

Kingma, D. P., & Welling, M. (2013). *Auto-Encoding Variational Bayes* (arXiv:1312.6114). arXiv. https://doi.org/10.48550/arXiv.1312.6114

Lee, S., Hoover, B., Strobelt, H., Wang, Z. J., Peng, S., Wright, A., Li, K., Park, H., Yang, H., & Chau, D. H. (2023). *Diffusion Explainer: Visual Explanation for Text-to-image Stable Diffusion* (arXiv:2305.03509). arXiv. https://doi.org/10.48550/arXiv.2305.03509

Leonardo.Ai. (n.d.). https://app.leonardo.ai/

LimeWire AI Studio. (n.d.). Retrieved April 19, 2024, from https://limewire.com/studio/image/create-image?model=blue-willow-v4&prompt=Create+a+bubble+diagram+for+an+architectural+project

Lin, K., Yang, Z., Li, L., Wang, J., & Wang, L. (2023). *DEsignBench: Exploring and Benchmarking DALL-E 3 for Imagining Visual Design* (arXiv:2310.15144). arXiv. https://doi.org/10.48550/arXiv.2310.15144

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. In D. Fleet, T. Pajdla,

B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision – ECCV 2014* (pp. 740–755). Springer International Publishing. https://doi.org/10.1007/978-3-319-10602-1_48

Liu, F., & Qian, Q. (2022). Cost-Sensitive Variational Autoencoding Classifier for Imbalanced Data Classification. *Algorithms*, *15*, 139. https://doi.org/10.3390/a15050139

Liu, J. (2023). Review of variational autoencoders model. *Applied and Computational Engineering*, *4*, 588–596. https://doi.org/10.54254/2755-2721/4/2023328

Liu, V., & Chilton, L. B. (2023). *Design Guidelines for Prompt Engineering Text-to-Image Generative Models* (arXiv:2109.06977). arXiv. https://doi.org/10.48550/arXiv.2109.06977

Lockard, W. K. (1982). *Design drawing*. Tucson, Ariz. : Pepper Pub. http://archive.org/details/designdrawing00lock

Loder, D. (2023, June 7). *Machines Imagining Interior Images: Investigating the significance of algorithmic 'text-to-image' processes for the practice and pedagogy of interior design*. Creative Curriculum: Supporting Creative Practice and Practitioners for the 21st Century. GSA Learning & Teaching Conference 2023, The Glasgow School of Art, UK. https://radar.gsa.ac.uk/9025/

LookX AI Cloud. (n.d.). https://www.lookx.ai/pc/gc

Mansimov, E., Parisotto, E., Ba, J. L., & Salakhutdinov, R. (2015, November 9). *Generating Images from Captions with Attention*. http://arxiv.org/abs/1511.02793

Marcus, G., Davis, E., & Aaronson, S. (2022). *A very preliminary analysis of DALL-E 2* (arXiv:2204.13807). arXiv. https://doi.org/10.48550/arXiv.2204.13807

Mijwil, M. (2015). *History of Artificial Intelligence* (p. 8). https://doi.org/10.13140/RG.2.2.16418.15046

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 (arXiv:2112.10741). arXiv. https://doi.org/10.48550/arXiv.2112.10741

O'Connor, R. (2023, September 29). *How DALL-E 2 Actually Works*. News, Tutorials, AI Research. https://www.assemblyai.com/blog/how-dall-e-2-actually-works/

Olaoye, F., & Potter, K. (2024). Deep Learning Algorithms and Applications. *Dissolution Technologies*.

OpenAI. (2023). *DALL.E 3 System Card*. https://openai.com/research/dall-e-3-system-card

Oppenlaender, J. (2023). A taxonomy of prompt modifiers for text-to-image generation. *Behaviour & Information Technology*, *0*(0), 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.Org*. https://arxiv.org/abs/2303.13534v2

Paananen, V., Oppenlaender, J., & Visuri, A. (2023). Using Text-to-Image Generation for Architectural Design Ideation. *International Journal of Architectural Computing*, 14780771231222783. https://doi.org/10.1177/14780771231222783

parsons, guy. (2022, July 22). *The DALL·E 2 prompt book*. Dallery.Gallery. https://pitch.com/v/DALL-E-prompt-book-v1-tmd33y/9a20735b-b010-4a6e-87c3-28852943087b

Pavlichenko, N., & Ustalov, D. (2023). Best Prompts for Text-to-Image Models and How to Find Them. *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2067–2071. https://doi.org/10.1145/3539618.3592000

Ploennigs, J., & Berger, M. (2023a). *Analysing the usage of AI art tools for architecture*. *4*, 0–0. https://doi.org/10.35490/EC3.2023.253

Ploennigs, J., & Berger, M. (2023b). *Diffusion Models for Computational Design at the Example of Floor Plans* (arXiv:2307.02511). arXiv. https://doi.org/10.48550/arXiv.2307.02511

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning Transferable Visual Models From Natural Language Supervision. *Proceedings of the 38th International Conference on Machine Learning*, 8748–8763. https://proceedings.mlr.press/v139/radford21a.html

Radhakrishnan, M. (2023). *Is Midjourney-Ai the New Anti-Hero of Architectural Imagery & Creativity? 11*, 94–114. https://doi.org/10.11216/gsj.2023.01.102270

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). *Hierarchical Text-Conditional Image Generation with CLIP Latents* (arXiv:2204.06125). arXiv. https://doi.org/10.48550/arXiv.2204.06125

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., & Sutskever, I. (2021). *Zero-Shot Text-to-Image Generation* (arXiv:2102.12092). arXiv. https://doi.org/10.48550/arXiv.2102.12092

Riahi, P. (2017). Expanding the boundaries of architectural representation. *The Journal of Architecture*, *22*(5), 815–824. https://doi.org/10.1080/13602365.2017.1351671

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022a). *High-Resolution Image Synthesis With Latent Diffusion Models*. 10684–10695. https://openaccess.thecvf.com/content/CVPR2022/html/Rombach_High-Resolution_Image_Synthesis_With_Latent_Diffusion_Models_CVPR_2022_paper.html

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022b). *High-Resolution Image Synthesis with Latent Diffusion Models* (arXiv:2112.10752). arXiv. https://doi.org/10.48550/arXiv.2112.10752

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Ghasemipour, S. K. S., Ayan, B. K., Mahdavi, S. S., Lopes, R. G., Salimans, T., Ho, J., Fleet, D. J., &

Norouzi, M. (2022). *Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding* (arXiv:2205.11487). arXiv. https://doi.org/10.48550/arXiv.2205.11487

Seneviratne, S., Senanayake, D., Rasnayaka, S., Vidanaarachchi, R., & Thompson, J. (2022). DALLE-URBAN: Capturing the urban design expertise of large text to image transformers. *2022 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, 1–9. https://doi.org/10.1109/DICTA56598.2022.10034603

Shaveta. (2023). A review on machine learning. *International Journal of Science and Research Archive*, *9*(1), Article 1. https://doi.org/10.30574/ijsra.2023.9.1.0410

Shojaee, S., Ali, S., Saremi, A., & Journal of Architecture and Urban Development, I. (2022). *Explaining the Methods of Architecture Representation Using Semiotic Analysis (Umberto Eco's Theory of Architecture Codes)*. *08*, 33–48.

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

Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., & Ganguli, S. (2015). *Deep Unsupervised Learning using Nonequilibrium Thermodynamics* (arXiv:1503.03585). arXiv. https://doi.org/10.48550/arXiv.1503.03585

Stable Diffusion launch announcement. (2022, August 10). Stability AI. https://stability.ai/blog/stable-diffusion-announcement

Tanugraha, S. (2023). A Review Using Artificial Intelligence-Generating Images: Exploring Material Ideas from MidJourney to Improve Vernacular Designs. *Journal of Artificial Intelligence in Architecture*, *2*(2), Article 2. https://doi.org/10.24002/jarina.v2i2.7537

Tsirikoglou, A., Eilertsen, G., & Unger, J. (2020). A Survey of Image Synthesis Methods for Visual Machine Learning. *Computer Graphics Forum*, *39*(6), 426–451. https://doi.org/10.1111/cgf.14047

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. ukasz, & Polosukhin, I. (2017). Attention is All you Need. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Wang, D., Ma, C., & Sun, S. (2023). Novel Paintings from the Latent Diffusion Model through Transfer Learning. *Applied Sciences*, *13*, 10379. https://doi.org/10.3390/app131810379

Wang, K., Gou, C., Duan, Y., Lin, Y., Zheng, X., & Wang, F.-Y. (2017). Generative adversarial networks: Introduction and outlook. *IEEE/CAA Journal of Automatica Sinica*, *4*(4), 588–598. https://doi.org/10.1109/JAS.2017.7510583

Wirz, F., & Gallo, G. (2020, January 1). The role of Artificial Intelligence in architectural design: Conversation with designers and researchers. *Proceedings of S.Arch 2020, the 7th International Conference on Architecture and Built Environment*. THE WAY IT'S MEAN TO BE. https://www.academia.edu/44902106/The_role_of_Artificial_Intelligence_in_architectural_design_conversation_with_designers_and_researchers

Xie, Y., Pan, Z., Ma, J., Jie, L., & Mei, Q. (2023). A Prompt Log Analysis of Text-to-Image Generation Systems. *Proceedings of the ACM Web Conference 2023*, 3892–3902. https://doi.org/10.1145/3543507.3587430

Xu, T., Zhang, P., Huang, Q., Zhang, H., Gan, Z., Huang, X., & He, X. (2018). AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1316–1324. https://doi.org/10.1109/CVPR.2018.00143

Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., & Yang, M.-H. (2023). *Diffusion Models: A Comprehensive Survey of Methods and Applications* (arXiv:2209.00796). arXiv. https://doi.org/10.48550/arXiv.2209.00796

Yildirim, E. (2022a). *Text-to-Image Generation A.I. in Architecture* (pp. 97–120).

Yildirim, E. (2022b, December 30). *TEXT TO IMAGE ARTIFICIAL INTELLIGENCE IN A BASIC DESIGN STUDIO: SPATIALIZATION FROM NOVEL*.

Yong, L., & Chibiao, H. (2022). A generative design method of building layout generated by path. *Applied Mathematics and Nonlinear Sciences, 7*. https://doi.org/10.2478/amns.2021.2.00168

Zhang, C., Zhang, C., Zhang, M., & Kweon, I. S. (2023). *Text-to-image Diffusion Models in Generative AI: A Survey* (arXiv:2303.07909). arXiv. https://doi.org/10.48550/arXiv.2303.07909

Zhang, H., Koh, J. Y., Baldridge, J., Lee, H., & Yang, Y. (2021). *Cross-Modal Contrastive Learning for Text-to-Image Generation*. 833–842. https://openaccess.thecvf.com/content/CVPR2021/html/Zhang_Cross-Modal_Contrastive_Learning_for_Text-to-Image_Generation_CVPR_2021_paper.html

Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., & Metaxas, D. (2017). StackGAN: Text to Photo-Realistic Image Synthesis with Stacked Generative Adversarial Networks. *2017 IEEE International Conference on Computer Vision (ICCV)*, 5908–5916. https://doi.org/10.1109/ICCV.2017.629

Zhang, Z., Fort, J. M., & Giménez Mateu, L. (2023). Exploring the Potential of Artificial Intelligence as a Tool for Architectural Design: A Perception Study Using Gaudí'sWorks. *Buildings*, *13*(7), Article 7. https://doi.org/10.3390/buildings13071863

Zhou, R., Jiang, C., & Xu, Q. (2021). A survey on generative adversarial network-based text-to-image synthesis. *Neurocomputing*, *451*, 316–336. https://doi.org/10.1016/j.neucom.2021.04.069

Zhuhadar, L. P. (2023). *A comparative view of AI, Machine Learning, Deep Learning, and Generative AI. Through this comparative lens, the diagram illuminates the distinct features as well as their overlapping facets between these fields. This comparison shines a light on their unique characteristics and shared elements, enabling us to appreciate the interconnectedness and individuality of these concepts. This*

understanding is crucial in realizing the full potential of AI and its multifaceted aspects in current and future applications. Own work. https://commons.wikimedia.org/wiki/File:Unraveling_AI_Complexity_-_A_Comparative_View_of_AI,_Machine_Learning,_Deep_Learning,_and_Generative_AI.jpg

APPENDIX

A.1 THE GENERATED IMAGES

Figure A.1: The generated images for the bubble diagram.

Figure A.2: The generated images for the conceptual sketches.

Figure A.3: The generated images for the floor plan drawings.

Figure A.4: The generated images for the cross-section drawing.

Figure A.5: The generated images for the elevation drawing.

Figure A.6: The generated images for the exploded view diagram.

Figure A.7: The generated images for the rendered 3D model.

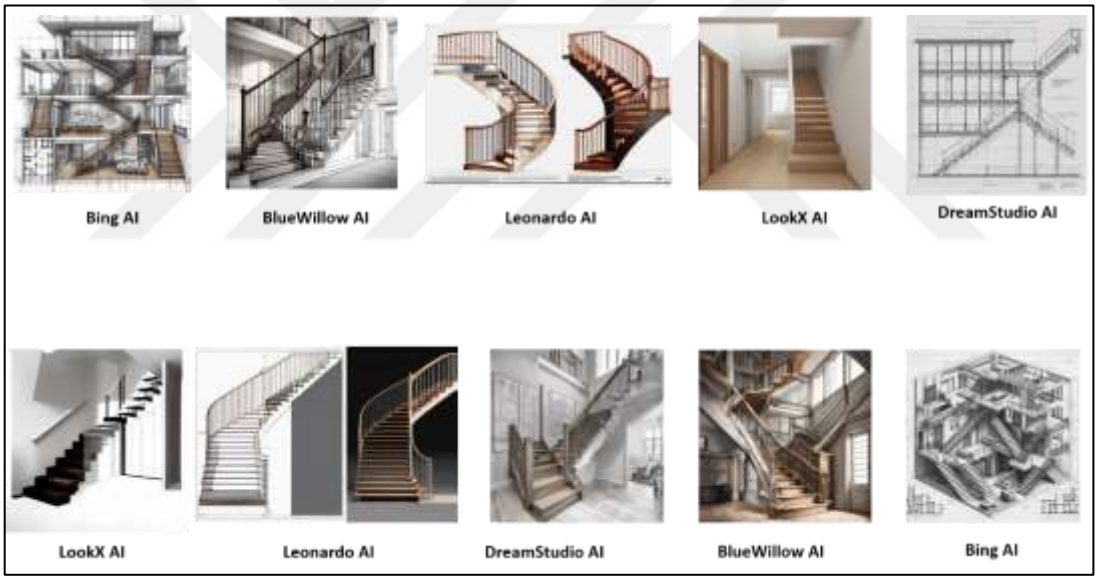


Figure A.8: The generated images for the technical detailed staircase drawings.

A.2 THE SURVEY DATA

Table A.1: The average scores for each representational technique.

Architectural representations	Average score
Bubble diagram	1,98
Conceptual sketches	2,71
Floor plan drawings	3,25
Cross section drawing	2,69
Elevation drawing	2,88
Rendered 3d model	3,30
Exploded view diagram	2,67
Architectural element details drawings	2,34

Table A.2: The average scores for each tool (continuous).

The architectural technique	The tools	The average Score
Bubble diagram	Bing AI	1,82
	BlueWillow AI	2,06
	Leonardo AI	2,34
	LookX AI	1,78
	Dream studio AI	1,91
Conceptual sketches	Bing AI	2.77
	BlueWillow AI	2.65
	Leonardo AI	3.01
	LookX AI	2.25
	Dream studio AI	2.89
Floor plan	Bing AI	2.45
	BlueWillow AI	2.97
	Leonardo AI	3.25
	LookX AI	3.88
	Dream studio AI	3.68
Cross-section	Bing AI	3.62
	BlueWillow AI	2.82
	Leonardo AI	2.98
	LookX AI	1.31
	Dream studio AI	2.73

Table A.2: The average scores for each tool.

The architectural technique	The tools	The average Score
Elevation	Bing AI	1.78
	BlueWillow AI	3.64
	Leonardo AI	2.66
	LookX AI	2.60
	Dream studio AI	3.73
Exploded view diagram	Bing AI	3.22
	BlueWillow AI	3.23
	Leonardo AI	2.49
	LookX AI	1.81
	Dream studio AI	2.60
Rendered 3D model	Bing AI	3.37
	BlueWillow AI	2.96
	Leonardo AI	3.67
	LookX AI	3.06
	Dream studio AI	3.47
Architectural element details drawings	Bing AI	2.15
	BlueWillow AI	2.06
	Leonardo AI	2.61
	LookX AI	2.07
	Dream studio AI	2.82

Table A.3: The average score of each image generated by each tool for the bubble diagram.

Bubble diagram	The tools	The average score for each image (the two images) generated per each tool	
	Bing AI	1,69	1,95
	Bluewillow AI	1,98	2,14
	Leonardo AI	2,17	2,50
	LookX AI	1,69	1,86
	DreamStudio AI	1,86	1,95

Table A.4: The average score of each image generated by each tool for the conceptual sketches.

Conceptual sketches	The tools	The average score for each image (the two images) generated per each tool	
	Bing AI	2,74	2,79
	Bluewillow AI	2,50	2,79
	Leonardo AI	3,00	3,02
	LookX AI	2,17	2,33
	DreamStudio AI	2,64	3,14

Table A.5: The average score of each image generated by each tool for the floor plan drawings.

Floor plan drawings	The tools	The average score for each image (the two images) generated per each tool	
	Bing AI	2,38	2,52
	Bluewillow AI	2,95	2,98
	Leonardo AI	2,93	3,57
	LookX AI	3,50	4,26
	DreamStudio AI	3,48	3,88

Table A.6: The average score of each image generated by each tool for the elevation drawing.

Elevation drawing	The tools	The average score for each image (the two images) generated per each tool	
	Bing AI	1,69	1,86
	Bluewillow AI	3,52	3,76
	Leonardo AI	2,07	3,24
	LookX AI	1,69	3,50
	DreamStudio AI	3,60	3,86

Table A.7: The average score of each image generated by each tool for the cross-section drawing.

Cross-section drawing	The tools	The average score for each image (the two images) generated per each tool	
	Bing AI	3,31	3,93
	Bluewillow AI	2,45	3,19
	Leonardo AI	2,48	3,48
	LookX AI	1,29	1,33
	DreamStudio AI	2,26	3,19

Table A.8: The average score of each image generated by each tool for the exploded view diagram.

Exploded view diagram	The tools	The average score for each image (the two images) generated per each tool	
	Bing AI	3,10	3,33
	Bluewillow AI	2,93	3,52
	Leonardo AI	2,48	2,50
	LookX AI	1,36	2,26
	DreamStudio AI	2,50	2,69

Table A.9: The average score of each image generated by each tool for the rendered 3D model.

Rendered 3D model	The tools	The average score for each image (the two images) generated per each tool	
	Bing AI	3,02	3,71
	Bluewillow AI	2,81	3,10
	Leonardo AI	3,45	3,88
	LookX AI	3,02	3,10
	DreamStudio AI	3,45	3,48

Table A.10: The average score of each image generated by each tool for the architectural element drawings.

Architectural element details drawings	The tools	The average score for each image (the two images) generated per each tool	
	Bing AI	1,74	2,55
	Bluewillow AI	1,93	2,19
	Leonardo AI	2,52	2,69
	LookX AI	1,90	2,24
	DreamStudio AI	1,95	3,69