EXPLAINABILITY OF GENERATIVE AI FOR ARCHITECTURE: INVESTIGATING THE ALIGNMENT WITH DESIGN INTENTIONS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MUSTAFA SİNA KARAOĞLU

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
BUILDING SCIENCE IN ARCHITECTURE

SEPTEMBER 2023

Approval of the thesis:

EXPLAINABILITY OF GENERATIVE AI FOR ARCHITECTURE: INVESTIGATING THE ALIGNMENT WITH DESIGN INTENTIONS

submitted by **MUSTAFA SİNA KARAOĞLU** in partial fulfillment of the requirements for the degree of **Master of Science in Building Science in Architecture, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** ________________

Prof. Dr. Fatma Cânâ Bilsel
Head of the Department, **Architecture** ________________

Prof. Dr. Arzu Gönenç Sorguç
Supervisor, **Architecture, METU** ________________

Examining Committee Members:

Prof. Dr. Ali Murat Tanyer
Architecture, METU ________________

Prof. Dr. Arzu Gönenç Sorguç
Architecture, METU ________________

Assoc. Prof. Dr. Mehmet Koray Pekeriçli
Architecture, METU ________________

Prof. Dr. Meryem Birgül Çolakoğlu
Architecture, ITU ________________

Prof. Dr. Mine Özkar Kabakçıoğlu
Architecture, ITU ________________

Date: 08.09.2023

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Mustafa Sina Karaoğlu

Signature :

ABSTRACT

EXPLAINABILITY OF GENERATIVE AI FOR ARCHITECTURE: INVESTIGATING THE ALIGNMENT WITH DESIGN INTENTIONS

Karaoğlu, Mustafa Sina
Master of Science, Building Science in Architecture
Supervisor : Prof. Dr. Arzu Gönenç Sorguç

September 2023, 97 pages

Generative AI systems have the potential to elevate computational design tools from mere design artifact representation aids to influential collaborators in the design process. However, due to the black-box nature of generative AI, it is challenging for users to grasp its mechanisms and ascertain if the system aligns with their design intentions, thereby limiting its full potential in the architectural design process. Addressing this issue, this thesis presents a specialized text-to-image generative model. Unlike off-the-shelf models, this model benefits from curated and structured training data, facilitating controlled image generation and systematic evaluation. The model is developed by fine-tuning a pre-trained Stable Diffusion model using 85,000 architectural images and structured textual labels, representing diverse architectural design concepts related to the images, extracted from architectural project descriptions via neural topic modeling. Through investigating how well these labels are represented in the generated outputs across various architectural design use cases, the study demonstrated the model's alignment with design intentions.

Keywords: Deep Generative Learning for Architecture, Architectural Design Process, Explainable Generative AI, Diffusion Models, Large Language Models

ÖZ

MİMARİ İÇİN ÜRETKEN YAPAY ZEKANIN AÇIKLANABİLİRLİĞİ: TASARIM AMAÇLARIYLA UYUMUNUN İNCELENMESİ

Karaoğlu, Mustafa Sina
Yüksek Lisans, Yapı Bilimleri, Mimarlık
Tez Yöneticisi: Prof. Dr. Arzu Gönenç Sorguç

Eylül 2023, 97 sayfa

Üretken yapay zeka (YZ), bilgisayarlı tasarım araçlarını sadece tasarım nesnesinin temsiline yardımcı olan araçlardan, tasarıma yön veren iş birlikçilere dönüştürme potansiyeline sahiptir. Ancak, üretken yapay zekanın ‘kapalı kutu’ yapısı, kullanıcıların bu sistemleri anlamasını ve tasarım amaçlarıyla uyumlu olup olmadığını belirlemesini zorlaştırmaktadır. Bu da, YZ’nin mimarideki kullanımını kısıtlamaktadır. Bu çalışmada, bahsedilen sorun, metinden imaj üreten bir YZ’nin özelleştirilmesi ve incelenmesi ile ele alınmıştır. Hazır modellerden farklı olarak bu model, seçilmiş ve yapılandırılmış mimari veriyle eğitildiğinden kontrollü imaj üretimine ve sistematik değerlendirmeye olanak tanımaktadır. Model bir ‘Stable Diffusion’ modelinin, 85.000 mimari fotoğraf ve bunlarla ilişkili mimari tasarım kavramlarını temsil eden yapılandırılmış metin etiketleri eşliğinde tekrar eğitilmesi ile oluşturulmuştur. Bu etiketler, mimari proje açıklamalarından, yapay sinir ağlarına dayanan bir konu modellemesi tekniği ile çıkarılmıştır. Bu model, mimari tasarım senaryoları ile test edilerek, mimari tasarım amaçlarıyla uyumu incelenmiştir.

Anahtar Kelimeler: Mimarlıkta Üretken Derin Öğrenme, Mimari Tasarım Süreci, Açıklanabilir Üretken YZ, Difüzyon ve Geniş Dil Modelleri

to human knowledge, from which machines learn

ACKNOWLEDGMENTS

First and foremost, I would like to express my gratitude to my supervisor, Prof. Dr. Arzu Gönenç Sorguç, for her guidance and understanding. She provided clarity and helped me to find my perspective in this emergent field of AI, even when there was no solid ground or established concepts to navigate.

My appreciation extends to the examining committee members Prof. Dr. Ali Murat Tanyer, Assoc. Prof. Dr. Mehmet Koray Pekeriçli, Prof. Dr. Meryem Birgül Çolakoğlu, and Prof. Dr. Mine Özkar Kabakçıoğlu. Their invaluable feedback and criticisms significantly influenced the course of my research.

A special thanks to the Digital Design Studio team, for creating an enriching environment that sparked enthusiasm and fostered meaningful discussions among students.

I would also like to thank my former colleagues, Muhammet Turşak, Ertem Yazıcıoğlu, Peter Mulvey, and Shayne Paterson for sharing their exceptional software development skills with me and aiding me in overcoming many technical challenges.

Words cannot express my gratitude to my parents, Turan Karaoğlu and Serpil Karaoğlu, for making every accomplishment in my life possible. I am also deeply grateful to my beloved partner Ege Naz Özer for her support and patience. Her encouragement in my scientific explorations and her timely reminders to narrow my focus have been invaluable. Lastly, I would like to thank my dog, Sıpa, who never left my side during countless hours of coding and writing.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

The rapid advancements in Natural Language Processing and Generative AI systems hold the potential to elevate computational design tools from mere design artifact representation aids to influential collaborators that can augment the design process. However, the artificial neural network techniques behind generative AI systems are inherently opaque, lacking a structure that allows humans to grasp clear insights into the model's capabilities and biases about architectural concepts. Consequently, it is challenging for users to ascertain if the AI system aligns with their design intentions and to harness its full potential for insightful design space exploration and better design decisions. While explainability methods have effectively tackled the challenges of more specific machine learning tasks, generative systems prove more elusive due to the absence of a formalized problem definition.

By framing the issues associated with generative models as design challenges, this research investigates the constraints and opportunities of generative AI within architectural design. Within this thesis, a distinct text-to-image diffusion model was built to explore use cases within architectural settings. This model is trained with semantically structured labels and it allows users to trace its training data back to specific architectural projects. Through this specialized model, systematic subjective evaluations were undertaken to examine its proficiency in accurately interpreting architectural intentions.

The model was trained with a dataset consist of 85,000 architectural photographs coupled with their textual descriptions from 11,000 single-family house projects sourced from the Archdaily online platform. The Low Rank Adaptation method was utilized to selectively adapt certain parameters of the existing Stable Diffusion model, turning it into a task-specific module. Rather than training the model directly with image-caption pairs, images were labeled in a structured manner, reflecting various architectural design elements. This label structure is derived from the architectural descriptions by utilizing sentence transformers, hierarchical text clustering, image clustering, and topic modeling techniques. Consequently, when this refined label system is employed, the model exclusively produces outputs grounded in the known projects and their specific design considerations.

1.1 Motivation

With the availability of vast amounts of data and the rapid advancements of machine learning algorithms, recent digital tools are becoming increasingly intuitive to interact with and capable of mimicking human language and reasoning. Today, AI systems can write code, generate lifelike artwork, conduct complex data analyses, understand and generate human-like text, assist in medical diagnosis, and predict patterns in vast datasets that are beyond human comprehension.

The motivation behind this study arises from the expansive possibilities AI introduces within the realm of architectural design. Historically, digital design tools served primarily as digitized extensions of pen and paper, enabling the representation of design artifacts. Progressing a step further, computational design methods have served for the representation of the rationale behind designs for the synthesis of physical artifacts. Yet, up until now, the agency predominantly resided with the designer and digital design tools haven't truly ventured into suggesting design alternatives or influencing a designer's focal point. Considering that the conceptual design phase, where a designer delineates primary design challenges, has an immense impact on the entire design process; introducing computers that enhance this stage with more insightful design decisions holds tremendous promise.

However, this emergent methodology is in its early stages, necessitating a deeper exploration and understanding.

The abovementioned capabilities of AI systems stem from a pivotal shift in programming methodologies. Traditional programming emphasizes explicit and clear rule definitions. In contrast, contemporary machine learning systems utilize a probabilistic approach. This paradigm shift allows us to surpass the inherent constraints of the human mind, which typically grasps systems through logical inferences and deductions. It grants developers the ability to derive rules directly from data, enabling them to integrate these into applications even without fully understanding the underlying mechanisms. This profound change has reshaped numerous aspects of life, pushing computational boundaries into the creative domain in unprecedented ways.

Yet, embracing these capabilities means yielding a part of our logical comprehension to algorithms, which raises concerns regarding trustworthiness. As we increasingly engage with these systems, they reshape our perspectives in more intricate and profound ways than traditional media ever did. Navigating the potential pitfalls while maximizing the transformative promise of AI necessitates a recalibration of our engagement patterns with such systems.

1.2 Problem Definition

Generative AI systems have begun to exert significant influence on creative design tasks. However, the generative capacities of these systems stem from their intricate neural network architectures, turning them into black boxes that operate in incomprehensible ways for human reasoning. While such models, especially text-to-image generative ones, can serve as invaluable design support tools, architects are constrained from fully harnessing their transformative potential in the design process due to these models' opacity.

While explainability techniques are being developed to address the challenges of neural networks, their primary focus remains on AI systems that possess explicit problem formalizations. In contrast, generative models lack a clear and easily testable objective, which makes it more difficult for users to have clear insights into the model's capabilities and biases about the architectural concepts.

As we transition into this new paradigm of interaction with computers, it becomes imperative for researchers, developers, and practitioners to adapt their modes of interaction with computers during the design process. However, due to the rapid pace of these advancements, there is a clear gap in understanding how to deploy generative AI for real-world design tasks, and how to develop affective ways of human-AI interaction for design.

1.3 Research Questions

Given the outlined problem definition, the following central questions arise. (1) Since generative models are trained on non-architecture-specific data for broader applications, to what extent can they discern and respond to the nuanced intentions and concepts embedded in architectural design? (2) Given that generative AI systems are not designed for a specific usage pattern, what potential roles or workflows can they augment the architectural design process? (3) Given the inherently opaque and incomprehensible structure of these models, how can users effectively assess and understand the capabilities and limitations of these models in practical design contexts?

1.4 Aim and Objectives

This research aims to demonstrate the implications presented by generative AI in architectural design and contribute to the broader understanding of their capabilities and limitations in the domain.

To achieve this aim, the objectives of the research are as follows: (1) addressing the issues of explainability and reliability by proposing a domain-specific text-to-image generative model and provide insights into its alignment architectural concepts, (2) exploring wide array of potential applications of generative AI in architectural design, including its utility in design research, creative exploration, aiding in problem resolution, and the formulation of final design artifacts. (3) developing methods or interfaces that facilitate architects and designers in comprehending the inner workings of these models, ensuring that users can make informed decisions by fully grasping both the strengths and limitations of the generative AI they deploy in practical design contexts.

1.5 Methodology

In light of the findings from the literature, a human-centric design approach is adopted, and the challenges associated with generative AI design tools were identified as design problems. As this approach requires understanding the specific needs of users, hypothetical use-cases for these tools within the architectural context were identified. The identified cases are addressed by using a uniquely structured text-to-image diffusion model. Finally, the results of the use cases are presented and subjectively interpreted by the author.

The developed model is a fine-tuned stable diffusion trained with 85,000 image-text pairs, representing 11,000 house projects on the Archdaily platform. The model was designed specifically to foster the identified design experiments. It stands apart with two characteristics. First, the texts that are paired with images in the training data are semantically structured labels, rather than unique sentences. The label structure is derived from textual descriptions using sentence embeddings, hierarchical clustering methods, and topic modeling. The aim of this is to facilitate systematic design experiments, which would be much more challenging with a completely unstructured model. Second, the model is trained on limited and known architectural projects. The aim of this is to make the input data that influenced a generated image traceable, in contrast with the off-the-shelf models' unknown training data that comes from the whole internet. Consequently, this model enabled a more structured visual output analysis based on the architectural relevance of the extracted visual patterns.

This model is used for a series of systematic subjective evaluations. Firstly, the coherence of the extracted design factors was interpreted to determine if a logical structure could be autonomously derived from the data without manual human intervention. Secondly, the fine-tuned stable diffusion model's ability to effectively learn from the training data is tested. It was imperative to ensure that the model had sufficiently segregated itself from its base training, rooted in billions of non-traceable image-text pairs. Finally, the architectural relevance of the generated images is assessed, ensuring that the architectural intent embedded in the prompts was accurately interpreted and captured by the model.

CHAPTER 2

BACKGROUND

In this chapter, a broad perspective of the AI research is presented. This includes the fundamental goals and challenges of the discipline, breakthroughs leading to the current state of AI, the reasons it became synonymous with artificial neural networks, and the episthemological implications of the shift from rule based approaches to todays probabilistic - connecitonist approach. Then, the AI applications in architecture were reviewed, with a particular focus on the assessment and evaluation of the generative models in design tasks.

2.1 Foundations of Artificial Intelligence Research

2.1.1 Mechanisation of Human Reasoning

One of the most perplexing challenges of AI research has been tackling tasks such as recognizing spoken words or identifying faces, that are instinctively simple and automatic for humans, yet remarkably complex to formalize (I. Goodfellow et al., 2017). As Nilsson (2010) highlights, the pioneers of the field were interested in "mimicking the higher levels of human thought", based on their limited introspection of human-like problem-solving that we do unconsciously.

Early dreams of creating a thinking machine can be traced back to modernity, when the cosmos was conceptualized as a machine. In this era, the mind emerged as a distinct entity from the body, as René Descartes elucidated in his Sixth Meditation (1641): "…I have a clear and distinct idea of myself, in so far as I am simply a thinking, nonextended thing; and on the other hand, I have a distinct idea of body, in so far as this is simply an extended, non-thinking thing…”. Descartes believed that humans are guided by an immaterial mind, whereas animals are governed by the laws of physics and lack reason. Thus, he believed that it is possible to construct automata that are indistinguishable from an animal but a machine attempting to imitate a human would be easily detected because it wouldn't use words and would lack general reasoning abilities. (Sharples et al., 1989) According to Sharples (1989), the modern AI research is still dealing with the same issues, language and general reasoning.

Progress in artificial intelligence research and development started with the represention of complex concepts and thoughts in a symbolic language (Nilsson, 2010). At the same era as Descartes, Gottfried Wilhelm Leibniz aimed to mechanize reasoning with a universal language, postulating that all knowledge could be distilled into simple propositions using lingua characteristica. His calculus ratiocinator would then verify the truth of these propositions (Nilsson, 2009, p. 29). Reflecting on this ambition, Leibniz stated, as cited by Davis (2018): “...how much better will it be to bring under mathematical laws human reasoning, the most excellent and useful thing we have.”. In the 19th century, George Boole's algebraic representation of logic paved the way for the logic gates in modern computers. Following Boole, Frege's Begriffsschrift, or "concept writing," allowed for the notation of propositions with their internal components. With these developments, the means to capture the intricacies of human thought in a machine-readable format began to take shape.

The ubiquitous presence of the term 'Artificial Intelligence' in contemporary dialogs owes much to the rapid shifts brought about by advancements in Neural Networks in the past decade, however, the field is established with this name in the early 1950s.

In his review, Nilsson (2009) highlights three major events in the mid-20th century, toward creating machines with human-like problem-solving abilities. "Session on Learning Machines" in 1955 unveiled several approaches such as using artificial constructs mimicking neurons for pattern recognition challenges or building a chess-playing program rooted in symbolic processing. The Dartmouth Summer Project in

1956, involving pioneers of the field like Arthur Samuel, Allen Newell, and Herbert Simon, was predicated on the belief that every aspect of learning and intelligence could be so precisely described that machines could be designed to simulate it. This event symbolizes the formal recognition of AI as a distinct research field. From the “Mechanisation of Thought Processes" symposium in 1958, Nilsson (2009) underscores several notable contributions from the pioneers. Minsky introduced heuristic programming and pattern recognition techniques. McCarthy presented an AI language system designed to enable machines to generate actionable insights. Oliver Selfridge introduced "Pandemonium", a model of cognition based on parallel processing in which 'demons' operated hierarchically to process and convey data. Significantly, Selfridge’s model could be adapted as either a neural network or a symbol-processing assembly, bridging two distinct AI paradigms.

2.1.2 Dealing with the Complexity and Uncertainity

Human reasoning often has to deal with incomplete and ambiguous information, which is challenging for logical systems. While humans can make to heuristic decision-making, drawing from intuition and past experiences, machines are predominantly built on analytical processes, prioritizing systematic and normative reasoning. This difference makes it inherently difficult for computers to navigate the complexities and uncertainties that humans easily address.

The complexity of a problem stems from the combinatorial explosion, where the volume of possible solutions can grow exponentially and make it computationally infeasible for machines to solve by testing all possible solutions (McCarthy, 1996; Minsky, n.d.; Nilsson, 2010). As illuminated by Minsky (n.d.), effective problem-solving requires techniques that employ prior knowledge to guide the search and make it more efficient, leveraging methods like "hill-climbing" which depend on understanding the structure of the search space. He also highlights the need for effective pattern recognition and classification of problems to decide on the most appropriate methods for solving them.

Arthur L Samuel’s work on game of checkers in the late 1950s is an illustrative example of strategies for simplifying and navigating through the game without exhaustively calculating every possible move. Despite being simpler than chess, checkers still presents 10^31 possible moves. Samuel employed the mini-max algorithm to craft a game tree from a given position, selecting the move with the maximum potential reward. Crucially, to reduce the vastness of possible game trees, Samuel utilized the alpha-beta pruning algorithm, which strategically truncated irrelevant tree branches based on heuristic evaluations.

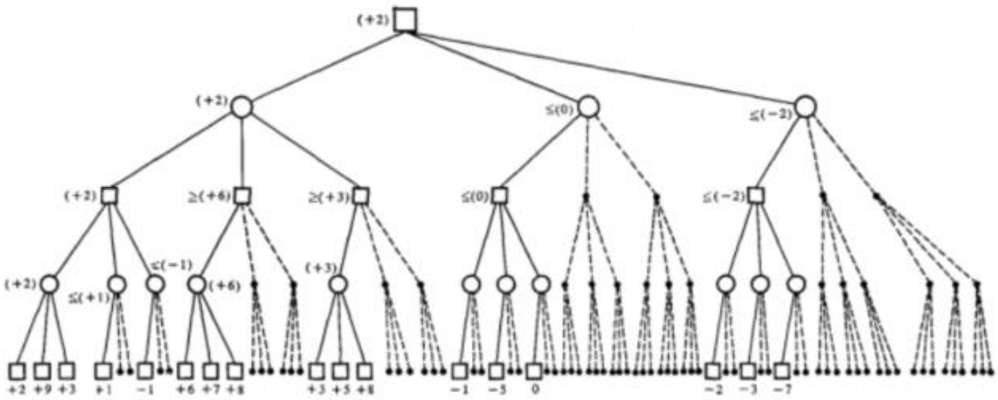


Figure 2.1 Game-tree (Samuel & Gabel, 2019)

In Figure 2.1, the first player shown as squares, the second player as circles. Each position is evaluated and credited by a set of parameters defined in Samuel’s work. Dashed lines are the terminated branches without further exploration and influencing the final. (Gabel, 2019)

Samuel’s work contributed with two pivotal methodologies: ‘rote learning’ and ‘learning by generalization’. These laid the foundation for modern machine learning, refining mappings devoid of direct programming. While 'rote learning' encapsulates the memorization of positions and their heuristic values for future reference, 'generalized learning' empowers the computer to compete against itself, progressively refining the evaluation functions of game positions (Gabel, 2019).

2.1.3 From Logicism to Connectionism

During the 1956 Dartmouth Summer Project, they intentionally chose the name 'Artificial Intelligence' to distinguish this new field from the ‘automata theory’ and 'cybernetics' (Nilsson, 2010). McCarthy (1996) argues that associating this new field with 'cybernetics', with its focus on analog feedback, could impose a limiting framework on the potential of artificial intelligence in areas like abstract reasoning, problem-solving, and knowledge representation. These researchers often called as ‘logicists’ or ‘symbolists’.

The approaches of Symbolism and Connectionism offer two distinct methodologies in artificial intelligence. Symbolism is grounded in structured thinking. It prioritizes clear knowledge constructs, logical reasoning, and systematic symbolic models to drive intelligent actions. On the contrary, Connectionism is inspired by observational learning theories. It focuses on discerning correlations and sequences directly from data, often with neural networks. (Goel, 2022) Historically, AI's symbol-oriented approach led to very specific, inflexible models designed for well-defined problems. These models, while successful within their narrow domains, struggled with exceptions, approximations, or heuristic knowledge, leading to limitations in broader contexts. On the other hand, the connectionist approach, represented by neural networks, tried to be universally flexible, but ended up making its own assumptions about similarities and inherent functionalities (Winston, 1990).

During the 1950s and 1960s, in pursuit of creating intelligent machines, AI researchers focused on identifying 'toy' problems to address. However, by the 1970s, AI began tackling real-life challenges such as natural language processing and computer vision. Yet, by the 1980s, the field experienced a loss of momentum and optimism. The AI research had to re-evaluate its promises in this period often referred to as the 'AI winter’ (Nilsson, 2010).

After the 'AI winter' of the 1980s, the perspective on neural networks, once considered as on the fringes (Nilsson, 2010) began to change. The growth in computational power and the rise of big data by the late 1990s provided the necessary boost. With these resources, the once computationally challenging deep neural networks started showing their strength, especially in tasks like image recognition and natural language processing. By the 2010s, neural networks had become central to today's AI.

2.1.4 Modern Artificial Neural Networks

Today, AI systems can emulate numerous human cognitive functions, fully automate certain tasks, and even outperform humans in some areas. It now plays a significant role in addressing various real-world challenges. (Nilsson, 2010) . These algorithms shape our interactions with web searches, content filtering on social networks, e-commerce recommendations, gadgets we use in everyday life, (Lecun et al., 2015), news feeds, financial transactions (Goodfellow et al., 2017), architectural design and visualization, structural analysis, material design and optimization (Baduge et al., 2022), More recently, owing to the large language models, AI systems are started to be used for computer programming, computational biology, medicine, creative work, law, social sciences (Kaddour et al., 2023)

This rise of AI occured with some milestones. The ascension of convolutional neural networks (CNN) enabled computers to discern and interpret intricate patterns and concepts that were previously deemed too abstract for anything but the human psyche. Generative adversarial nettowks (GAN) enabled computers to create. The transformer neural networks led unforeseen acceleration in natural language processing, transforming words and sentences from mere discrete entities to numerical vectors that allows us to capture semantic nuances.

While the ConvNets have been increased the success of image classification tasks, the generative ability of deep learning algorithms was limited until Generative Adversarial Networks developed.

A GAN consists of a generative model and a discriminator model trained simultaneously. The generator starts with generating a random noise. While the discriminator classifies the outputs of the generator as real or fake, the generator tries to maximize the probability of the discriminator making a mistake. This approach is achieved by training the system with backpropagation. While both discriminator and generator evolve together, the results become highly realistic. (Goodfellow et al. 2014) Over the past 8 years, GAN's are improved by many researchers. Each of them contributed to the algorithms by overcoming a challenge and creating a suitable model for a different purpose. Researchers also share their codes, so that others can use, modify, and improve their models.

While the first GAN of Goodfellow finally achieved image generation there was no control over the generation process. ConditionalGAN (Mirza, et al., 2014), contributed to GANs by conditioning the generator and discriminator on extra information such as class labels or multi-modal labels to be able to control the output.

ArtisticStyleTransfer (Gatsy, et al., 2015) and FastPhotoStyle (Li, et al., 2017) developed an algorithm that can learn the style of an image and transfer it to the context of another image. CycleGAN (Zhu, et al., 2017) and Pix2Pix (Isola, et al., 2017) are image-to-image translation models that allowed the application of a generic approach for different training tasks.

StyleGAN (Karras, et al., 2019) and StyleGAN2 (Karras, et al., 2020) from NVIDIA researchers, are powerful style-based conditional GANs that can generate realistic human faces that are indistinguishable from real. AdaGAN (Karras, et al., 2020) provided an adaptive discriminator augmentation mechanism to be able to train these models with limited data without increasing the bias. CreativeGAN (2021) contributed to GANs by assuring the novelty of generated images which has a bias to the training data.

The Transformer architecture, introduced by the "Attention is All You Need" (Vaswani et al., n.d.) paper, transformed the NLP landscape. Transformers replaced recurrent connections with self-attention mechanisms. It enabled better parallelization and allowed to capture long-range dependencies effectively. Models like GPT-3, T5, and BART have achieved state-of-the-art performance on tasks like language generation, machine translation, and text understanding.

2.1.5 Explainability

AI systems have become complex and sophisticated, making it difficult to understand how they reach their particular decisions or predictions. In some cases, even in principle, it is impossible to comprehend what happens in a neural network that led to a specific output. As AI systems increasingly take on decision-making tasks, this opacity raised concerns about their safety, reliability, accountability, fairness, and bias. These concerns are being addressed by an increasing volume of explainability research from multiple disciplines like human-computer interaction, computer science, design informatics, ethics, and law.

The 'black-box' models can create a lack of trust in AI which is particularly important in fields such as healthcare, finance, and law, where AI decision-making can have significant real-world implications. For example, if an algorithm used to make medical diagnoses makes a mistake that leads to harm to a patient, it can be difficult to determine who is responsible for the mistake. Similarly, if an algorithm used to determine loan eligibility produce discriminative results for minority groups unfair and unjust decisions can damage public trust. To address these challenges, in 2016, the European Union's General Data Protection Regulation included a "right to explanation" clause, which requires companies to provide explanations for the AI-made decisions that can "significantly affect" the subjects (Goodman & Flaxman, 2016).

The need for explainability does not only stem from the risks of AI, but also from human curiosity and the need for comprehension. When an opaque model is used to answer a scientific question, the findings are hidden from the researchers if there are no explanations but only predictions. In this case, the model itself becomes the source of scientific knowledge. (Molnar, 2022).

Carpo (2017) describes the impact of data richness by comparing the inductive and inferential methods of modern science to a hypothetical big-data scenario where every piece of data from an event is captured and available for retrieval. He characterizes the modern scientific method with its reliance on repeatedly observing an event, deriving patterns, and then strictly prioritizing them. This leads to generalizing mathematical formulas with the assumption that future events can be predicted by those formulas. In the hypothetical big data scenario, however, the reliance on such mathematical formulas could become obsolete. Instead, one might simply search a comprehensive database for a similar event under comparable conditions and directly observe the outcome. Yet, Carpo (2017) emphasizes that humans need these formulas while computers do not, and he criticizes the contemporary trend where theory-making is becoming undervalued, especially in a culture where much can be directly retrieved from data or computationally simulated.

While simpler models such as linear regression, logistic regression, and decision trees are intrinsically interpretable, more complex ‘black-box’ models such as neural networks require ‘post-hoc’ explanations after they are trained (Liao, n.d.). Guidotti et al. (2018) define three post-hoc explanation approaches. (1) Global model explanations focus on explaining the overall behavior of a model. This is often done by training an intrinsically interpretable to approximate the black-box model. (2) Outcome explanations focus on a particular outcome and investigate the importance of the features for the result of the model. (3) Caunterfactual inspection asks “why not” by changing an input until the outcome changes.

Disentanglement is a common approach to directly interpret the model's internal processes is creating mappings between high-dimensional inputs and low-dimensional representations to make the the model's latent dimensions more meaningful and easily examined by humans (Sun et al., 2022; Chen et al., 2018).

2.1.6 Explainability of Generative AI

The release of text-to-text and text-to-image generative models like DALL-E, Chat-GPT, Stable Diffusion, and Github Copilot in 2022 has led to a surge in the number of applications and their adoption. These recent models can perform well on a broader range of tasks while the traditional AI systems mostly perform on classification and regression problems (Maslej et al., 2023). However, this exciting capability of generative models comes with its bigger challenges regarding the ethical and safe use of them.

Utilizing larger general-purpose generative models can enhance creativity and diversity but it may lead to a compromise in the coherence and accuracy of the generated content. For example, conversational AI models like Chat-GPT create ‘hallucinations’ where the model confidently comes up with factually false outputs which makes it harder to rely on them for critical decisions (Maslej et al., 2023).

Explainable AI remains less clearly defined for generative AI compared to narrower discriminative tasks where significant progress has been made (Sun et al., 2022). Doshi-Velez & Kim (2017) suggest that the demand for explainability arises from the incomplete formalization of a problem.

These systems allows infinite use cases that can users come up with, which leads to subjective evaluation criteria like *‘you know it when you see it’*. Therefore, explainable generative AI is closer to an interaction design problem rather than an optimization problem. Sun et al. (2022) address this design problem by asking what knowledge users need to possess in order to effectively work with a generative AI model and achieve their goals.

In recent years, numerous AI applications reached end-users and explainability has become more than just a requirement for data scientists, developers or researchers to understand the models they create. This made the field embrace human-centric approaches which focus on how the user perceives the explanation. (Liao, n.d.) The Human-Centric Explainable AI framework places the question 'What do users need to understand about AI systems?' at its core, and explainability should not be limited to technical explanations of how the algorithm works (Sun et al., 2022)

2.2 Artificial Intelligence in Architectural Design

2.2.1 Modeling the Design Process for Computers

Architectural design, according to Harputlugil et al. (2014), “is an iterative process, having numerous parameters that are constantly evaluated through feedback”. According to Lawson (2005), as also cited by Harputlugil (2014), design problems cannot be comprehensively stated, since many aspects of these problems do not surface until some attempts at generating solutions have been made. The author suggests that the design problems and their corresponding solutions are in a state of continuous ‘dynamic tension’, as the objectives and priorities can shift during the process as the implications of potential solutions begin to unravel.

The term design serves as both a verb that refers to the process, and a noun that refers to the artifact (Lawson, 2005; Dino, 2012). When viewed as a process, design is inherently complex, a fact that is supported by an extensive volume of literature. Chatzikonstantinou, et. al (2017) explain the source of complexity with ‘combinatorial explosion’ and ‘non-linear relationships between the object properties and their abstractions’.

The designers, however, commonly navigates this complexity and make these design decisions by leaning on their intuition and professional experience, rather than explicitly modeling the design problem. In their studies on explainability, Wang et al., (2019) and Bielski et al., (2022) explain this two distinct decision-making strategies with the dual-process theory, a commonly adopted approach in cognitive psychology research on human judgments. (Croskerry, 2009; Evans, 1984). According to Croskerry (2009), there are two fundamental decision-making approaches: intuitive (heuristic) and analytical (systematic, normative). Intuition heavily relies on the designer's past experiences and it involves pattern recognition and gestalt effects and to inform decision-making. With this approach, decisions are typically taken under conditions of uncertainty, rapidly, and with the help of heuristics or cognitive shortcuts. Preconscious cognitive processes also emerge within this framework, indicating that some level of perception and analysis might occur outside of conscious awareness. Contrastingly, the analytical approach reflects a systematic method of decision-making which typically arises when uncertainty is low. The decisions made with this approach are more closely aligned with normative reasoning and rationality.

Chatzikonstantinou & Sariyildiz (2017) defines two distinct design criteria. The first one is “suitability”, which is the performance of the solution in terms of meeting the design goals and objectives. The second is “desirability”, which is the design preferences at the lowest abstraction level which are not related to the design goals. The authors state that the solutions that meet both criteria is challenging and computational decision support tools are efficient to handle the complexity of this. These methods are recently started to be applied for architectural design tasks, while they were in use for engineering tasks.

2.2.2 Evolution of Computer Aided Design

The early Computer Aided Architectural Design (CAAD) research started in the 1960’s. According to Koutamanis (2005), a small volume of publications from this era mainly discussed the technological advancements, architectural application of these technologies, theoretical work in the domain of architecture, and the exploration of related fields like AI. Specifically, the research on the technological foundations is mainly focused on computer graphics for automating the drafting tasks and the computational representation of design problems. While the CAAD became a popular and known research field in 1970s due to the availability of comprehensive reviews, it became a recognized and coherent area in the 1980s when different parties such as CAAD theorists and computer drafting practitioners have gained better understanding of the whole spectrum of aspects of computer-aided activities. (Koutamanis, 2005) When it comes to the 1990's CAAD became a mainstream field with its dedicated conferences and journals and dramatically increased volume of publications. However, according to Koutamanis (2005), early applications in the 1990's were limited to transferring analog processes to the computer, and this computerisation process of architectural design practice was mostly due to the democratization of computers, rather than the intellectual developments in CAAD.

A similar distinction to 'Computer-aided drawing and Design Automation' have been made by Gero (1994) to define the main areas in the field of computer-aided design. As cited by Dino (2012): "the representation and production of the geometry and topology of designed objects" and "the representation and use of knowledge to support or carry the synthesis of designs". While the first one is related to the widely-adopted CAD tools, the latter led to the development of generative approaches that uses computation for the design process. (Dino, 2012)

2.2.3 Generative Adversarial Networks (GAN)

Generative Adversarial Networks (GANs), first introduced in 2014, have quickly become a prominent tool in the art and design field. They marked the first time machines could recognize intricate visual patterns, such as styles, and generate new ones. The algorithm grabed attention of architects, since the field always had a significant emphasis on using imagery for inspiration, metaphor, and idea generation.

GAN's comprising both a generator and a discriminator model. The generator initiates the process with random noise, while the discriminator assesses these outputs for authenticity, the generator refines its creations. This iterative process leading to increasingly realistic results.

Over the years, the original GAN proposed by Goodfellow et al (2014) has been enhanced by numerous researchers, each adding layers of complexity and control. The advent of ConditionalGANs (Mirza & Osindero, 2014) brought controllability into the picture by conditioning the generator and discriminator on extra information such as class labels or multi-modal labels. Style Transfer (Li et al., 2017) made it possible to learn a style of an image and apply to another one, CycleGAN (Zhu et al., 2017) and Pix2Pix (Isola et al., 2017) added versatility for image-to-image translation. The more recent contributions, like NVIDIA's StyleGAN (Karras et al., 2020) series, have demonstrated the astonishing potential of GANs by crafting images so real that they blur the line between artificial and authentic.

One of the most prominent examples of the application of GANs in architectural design is ArchiGAN (Chaillou, 2020). This model levereges Pix2Pix HD algorithm developed by NVIDIA. The model is trained with building footprints of houses, the annotated layout by program, and furniture layout. The trained model takes the site shape as input and then progressively yielding a detailed architectural layout of the building.

However, despite the visual success of GANs, they could not have pivotal role in pragmatic design applications. The reasons are manifold: GANs, at their core, specialize in unconditional creations, image-to-image translations, or somewhat restrictive class-conditional productions. The challenge is navigating in the high-dimentional latent vector space, the realm where an image's granular details are distilled into lower-dimensional features. Conditioning a model to produce an image requires defining a vector within this latent space. Given the multidimensionality and complexity of this space, discerning which dimensions equate to which architectural elements becomes challenging. The present solutions are limited: either training the model on manually labeled data and then base image production on these labels, adapt the model to produce images influenced by input images, or interpolate between existing images. The latter, however, merely fuses existing images, rather than creationing novel, purposeful image which could be a part of practical architectural design process.

Figure 2.2The authors previous work on navigating in the latent vector space with vector arithmetics

2.2.4 Evaluation Methods of Generative Models

Neural networks, often described as 'black boxes' due to their complex inner workings, pose significant challenges in understanding and interpreting their processes. This intricacy is especially pronounced in generative models like diffusion models and Generative Adversarial Networks (GANs) which does not have a clear objective function. In the field, both quantitative and qualitative evaluation methods exist. While quantitative methods are crucial for the development of these models and addressing technical challenges, the alignment of the model with human needs and intentions often requires qualitative evaluations based on human observation.

Quantitative evaluation techniques comes with their own set of limitations. For instance, the Inception Score, which evaluates the diversity and quality of generated samples by examining the entropy of the output class distribution, can still yield high values for a model that memorizes a single example (mode collapse) for each ImageNet class. (Betzalel et al., 2022) Another method, Fréchet Inception Distance (FID) quantifies the quality of generated samples by comparing the Gaussian-distributed features of synthetic and real data using the Inception network, but its major limitation is its assumption that these features inherently follow a Gaussian distribution, which isn't always true in real-world datasets (Borji, 2018). FID score, according to Otani et al. (2023), often diverges from human perception. The author found that the CLIPscore method that evaluates image-text alignment is already saturated with the highly capable state-of-the-art text-to-image generative models.

As a result, in the face of these limitations, qualitative measurements based on human evaluations is an inherent need of the generative models since it involves human perception and understanding. However, there aren't established practices for human evaluation, leading to diverse protocols adopted by researchers. (Otani et al., 2023).

One of the main concerns in the evaluation of generative models is the relevance of visual characteristics encapsulated by the model. Huang et al. (2021) investigates this relevance in the context of architecture, where images have historically served as inspirations, metaphors, and analogies, as authors highlight. Their aim was to harness a combined intelligence of humans and AI to produce novel designs, ensuring that these designs resonate with culturaly and architecturaly relevant. The authors trained DCGAN and SAGAN algorithms with 6,000 images of Swiss mountain alpine refuges, that are selected based on the human subject's mental image of "Swissness." After GAN training, images are generated with latent space interpolation, then human subjects reviewed these generated images and identified those that aligned with their evolving concept of "Swissness." By marking specific images in the latent space, they guided the generation process, refining the images. To understand the architectural and cultural concepts tied to Swissness in the context of the images, the outputs are critically evaluated with textual analysis of the human interpretations and decomposition of architectural elements of generated images.

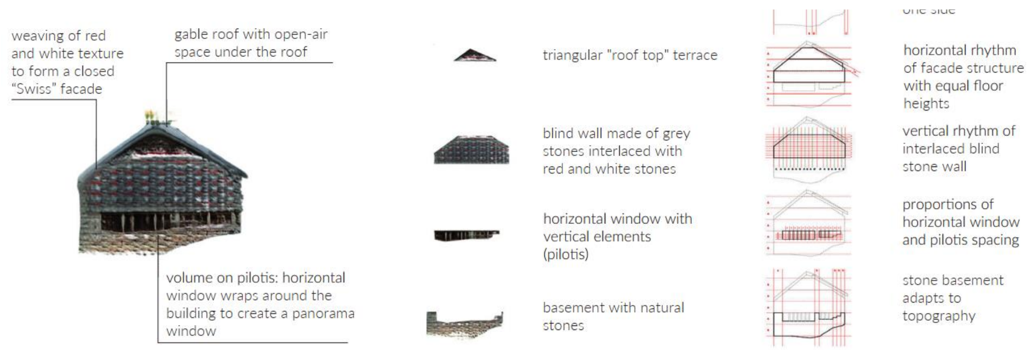


Figure 2.3 Huang et. al, (2021) human interpreatation of 'Swissness' in the generated images

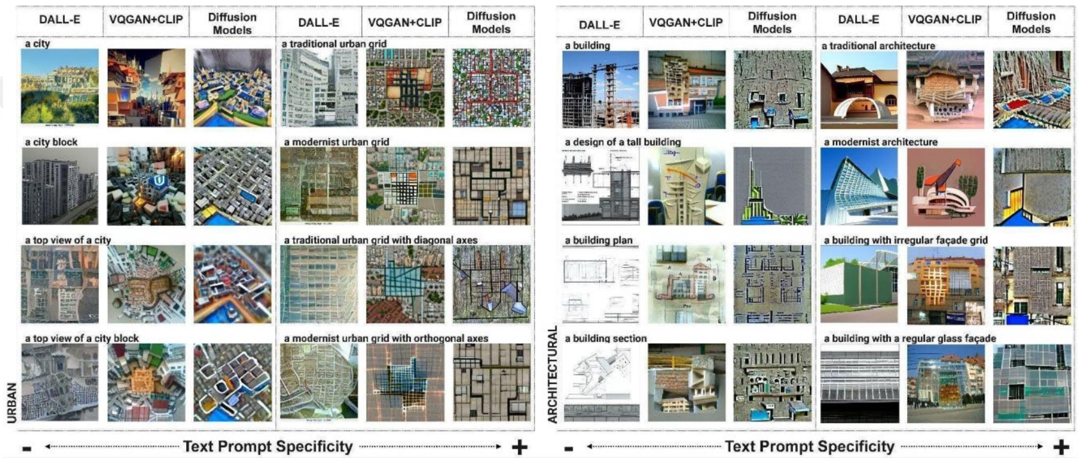


Figure 2.4 Experiment of Bolojan et. al, (2022) comparing the outputs of generative models when prompted with different specificity

In their 2022 study, Bolojan et al. investigated the application of text-to-image generative models in architectural design. They highlighted that textual descriptions offer a reductionist view of complex architectural concepts, and it is challenging to accurately bridge textual descriptions with their visual counterparts. In their first experiment, they compared the outputs of discreed models such as DALL-E, VQGAN/CLIP, and Diffusion models with varying levels of text prompt specificity. They observed that while some models like DALL-E were flexible, others required more precise prompts for effective results. Although hyperparameter fine-tuning improves the control of the behavior of models, authors highlighted the challenge of making language models that simultaneously focus on different abstraction levels involved in the design process. In the second experiment, Bolojan et al.

interconnected different domain-specific models that are curated by their strenghts in particular tasks. They used the output of one model as a reference image to guide the generation of another model which is specialized on another task. They found that this multi-modal and domain specific approach significantly improved architectural representations at various scales and allowed a more targeted and flexible conceptual space search.

CHAPTER 3

DEVELOPMENT OF THE PROPOSED MODEL

The proposed text-to image generative AI is built with a dataset comprising architectural photographs and corresponding textual descriptions of 11,000 single-family house projects fetched from the Archdaily online platform.

The model is constructed in two distinct phases. Firstly, architectural design factors are extracted from the textual descriptions through topic modeling techniques, which are complemented by image clustering and manual data preparation for geographical factors. Secondly, the Stable Diffusion model is fine-tuned using the extracted design factors as labels, employing the Low-Rank Adaptation method.

The Low-Rank Adaptation method is employed for selectively adapting certain parameters of a pre-trained Stable Diffusion model to create a task-specific module. Instead of directly training the model with image–caption pairs, the images are structurally labeled by a variety of architectural design factors. The label structure is derived from hierarchical text clustering that leverages sentence transformers, pre-trained language models, image clustering, and manually prepared geographical data. As a result, when the special label system is used, the model only generates the outputs based on the known projects and their known design considerations. This allows users to explore design options based on a variety of factors, such as volumetric composition, spatial perception, and architectural meaning. While allowing users to generate unique architectural images, the model also allows users to conduct formal analysis by accessing the actual data.

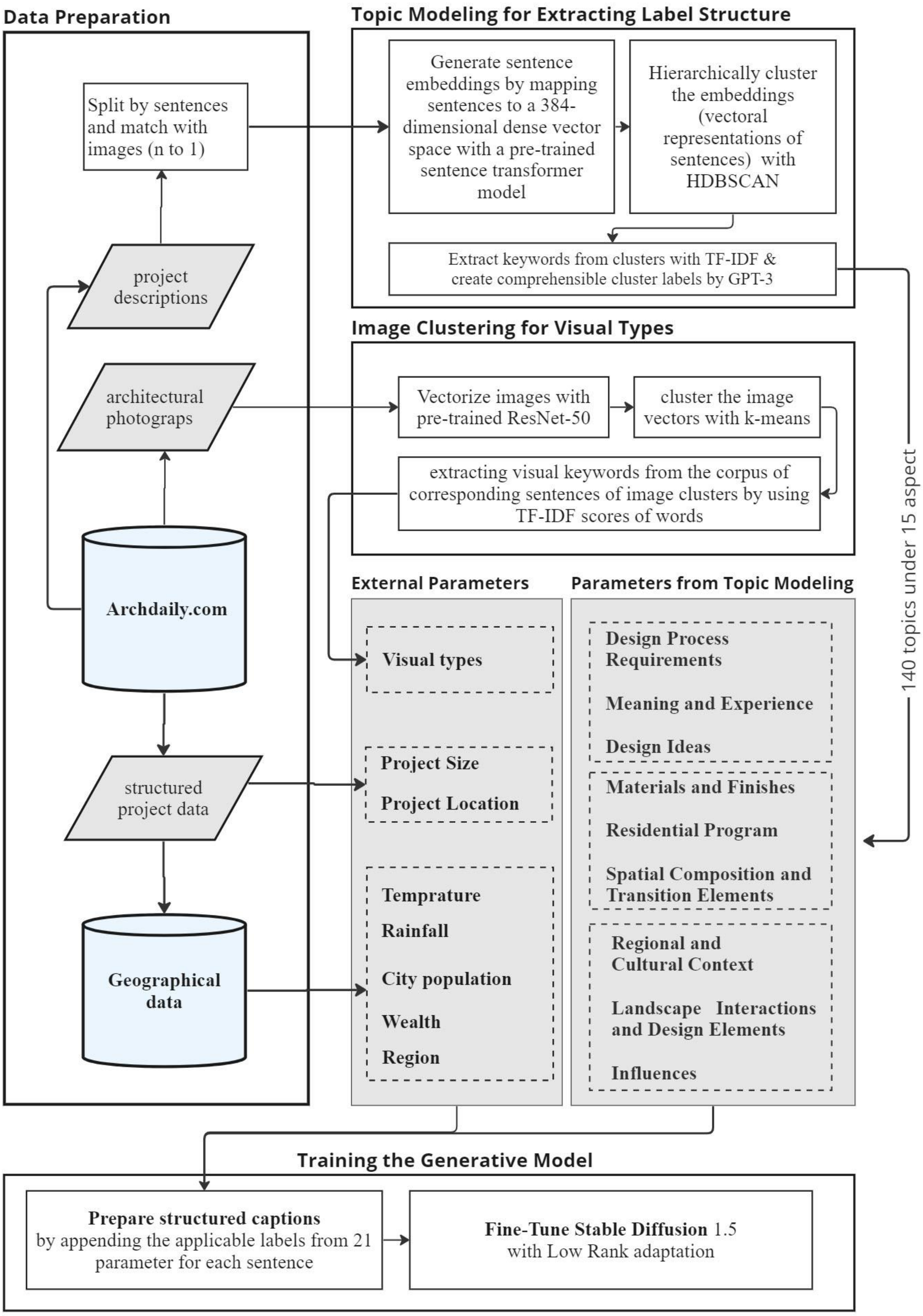


Figure 3.1: Pipeline of the Proposed Model

3.1 Data Collection and Preparation

The dataset consists of 11,300 residential buildings, containing a total of 170,000 sentences and 85,000 images. The design factor analysis is focused on house projects to achieve a more targeted and focused approach, allowing for a deeper understanding of the specific factors that influence the design and construction of residential buildings.

The project descriptions that are provided by the architects, and corresponding architectural photographs are downloaded from the Archdaily.com website by using a web crawling tool. To ensure the focus remains on visual patterns and accurate use for the training of the generative model, drawings, and diagrams are eliminated from the dataset. Unique identifiers are applied to each image, paragraph, and sentence to establish the relationship between the textual and visual components of each project.

Additional data, including average temperature, yearly rainfall, the population of the cities, and the size of the project, are manually added to the dataset. For easier handling and training, the continuous variables are discretized by the percentile-based discretization function of the Pandas library, and the countries are also grouped by sub-continents.

3.2 Potential Biases and Errors in Data Collection

It should be noted that there may be a low correlation between some of the sentence-image pairs due to issues during matching. The placement of images and sentences may depend on editorial preferences, causing the paired sentence to describe another image. Despite this potential margin of error, the distribution is based on a significant number of sentences (170,000), mitigating the risk of low accuracy in the distribution process. The accuracy of this process is further demonstrated through random collages; however, this margin of error should be taken into account.

The distribution of collected projects is not equal across all regions, as shown in the world map below. This bias should be considered when interpreting the model's outputs.

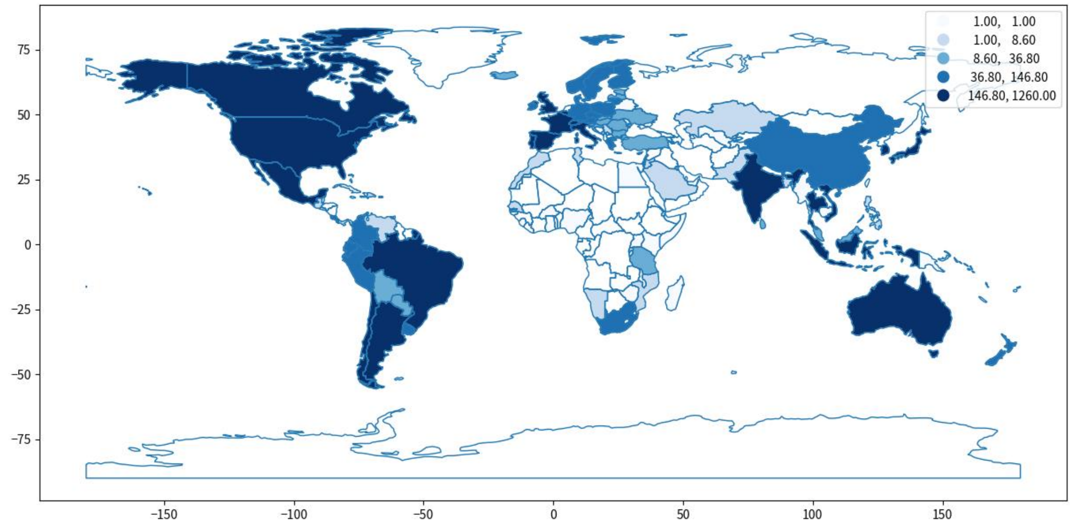


Figure 3.2: Distribution of Architectural Projects in the Dataset by Country

3.3 Neural Topic Modeling for Archıtectural Design Factor Analysis

The textual data is analyzed using the BERTopic model (Grootendorst, 2022), which is based on the BERT (Bidirectional Encoder Representations from Transformers) language representation model developed by Google Language AI (Devlin et al., 2018). BERT generates contextualized word and sentence representations, considering the meaning of the entire sentence rather than just individual words, as conventional topic modeling techniques do.

The topic modeling process involves several steps. First, document embeddings are generated using a pre-trained language model. Then, the dimensionality of the embeddings is reduced to create semantically similar clusters of documents. Finally, topic representations are extracted using a class-based version of TF-IDF.

The BERTopic model is implemented using the SentenceTransformer 'all-MiniLM-L6-v2' to generate document embeddings. The SentenceTransformer is a framework for computing dense vector representations of sentences, which are then used as input for the BERTopic model. The embeddings are 384-dimensional vectors that capture the semantic content of the sentences.

3.3.1 Deriving Semantic Similarities of Sentences

The SentenceTransformer model is used for deriving semantically meaningful sentence embeddings to perform the clustering. Sentence Transformer is an extension of the BERT model that is specifically tailored for generating semantically rich sentence embeddings (Reimers & Gurevych, 2019). It utilizes a pre-trained BERT network for feature extraction.

This pre-trained BERT model converts input text into contextual token embeddings, where each word is represented as a dense vector that encapsulates its context within the sentence. To create a single vector to represent the entire sentence rather than just individual tokens SBERT employs a pooling operation (like mean or max pooling) on the token embeddings. This operation amalgamates the individual token vectors into a single vector that represents the overall semantics of the sentence. The end result is a dense vector for each sentence, where the dimensionality is defined by the underlying BERT model. For 'all-MiniLM-L6-v2', this is a 384-dimensional vector space.

These dense vector representations, or embeddings, are arrays of floating point numbers, the size of which is determined by the dimensionality of the vector space. For instance, with 'all-MiniLM-L6-v2', each sentence is translated into an array of 384 numbers. The relative distances between these vectors in the embedding space correlate with the semantic similarity between sentences. That is, sentences with similar meanings are placed close together in this space, while dissimilar sentences are further apart.

Consequently, the SentenceTransformer model significantly simplifies and accelerates the process of semantic search, clustering, and information retrieval in large text corpora. Instead of having to compare each pair of sentences individually, which is a computationally expensive task, we can simply measure distances in the embedding space, a process that is much more efficient and scalable.

3.3.2 Hierarchical Clustering from Sentence Embeddings

To identify the themes in architectural descriptions, the sentence embeddings that captures semantic relations are clustered in two steps. UMAP (Uniform Manifold Approximation and Projection) algorithm is used for the dimensionality reduction of the sentence embeddings, and HDBSCAN algorithm is for clustering the reduced embeddings. These techniques are also recommended by the creators of BERTopic (Grootendorst, 2022), ensuring their compatibility with the approach of this research.

The transition from a high-dimensional space, such as the one created by 'all-MiniLM-L6-v2' SentenceTransformer model, to a lower-dimensional one that is suitable for visualization and interpretation is accomplished through UMAP. This method is known for preserving both local and global characteristics of high-dimensional data. It is similar to t-SNE but is computationally more efficient and better suited to data where clusters are not clearly separated. In this context, UMAP serves as a valuable tool for quickly reviewing the quality of document clusters during hyperparameter tuning. (McInnes et al., 2018)

After the dimensionality reduction, HDBSCAN, a density-based clustering technique, is employed to discover patterns and themes in the data. This algorithm stands out due to its capacity to explore varying levels of granularity, without the need for a pre-defined number of clusters. Furthermore, it is capable of recognizing clusters of diverse shapes and sizes, and it doesn't enforce every document into a cluster, thus effectively handling outliers.

However, the HDBSCAN produces a significant number of outliers. To address this, the outliers are reduced by calculating the cosine similarity of embeddings. This method measures the cosine of the angle between two vectors, providing a measure of how similar they are. If the cosine similarity is close to 1, the documents are very similar; if it's close to -1, they're very dissimilar.

Ultimately, the selection of the pre-trained model and the fine-tuning of the hyperparameters for clustering, dimensionality reduction, and outlier reduction are established through manual iterations and reviews, based on visualization results at each stage. This process results in the grouping of document clusters (topics) under parent topics (aspects), which is a suitable structure for architectural design factor analysis and the training of generative models.

showing the topics that are grouped under the 'Materials and textures' aspect in yellow. The dendrograms that vşsualize the hierarchical relations of the topics are based on the c-TF-IDF scores of the clusters, instead of the cosine similarity of the embeddings. Therefore, UMAP visualizations and the dendrograms prove the same grouping independently.

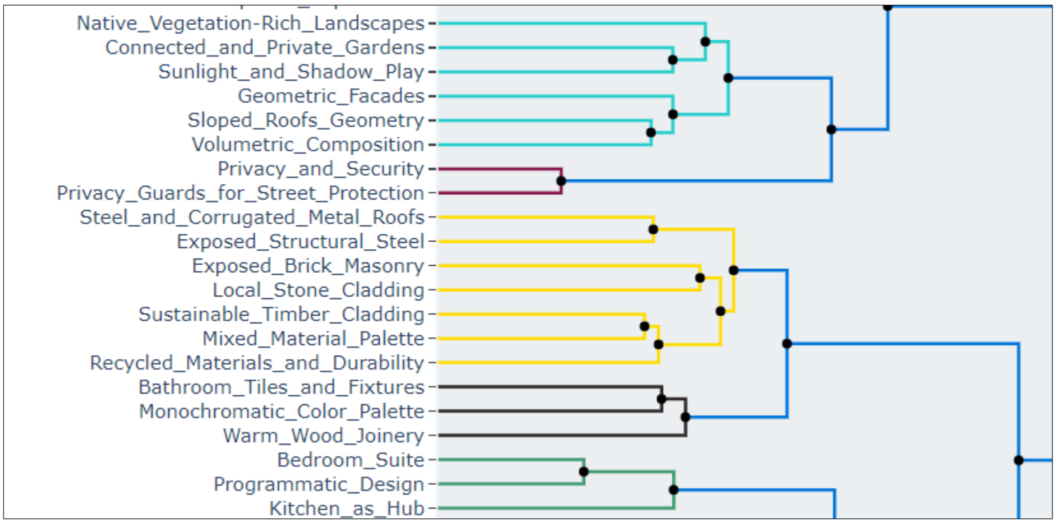


Figure 3.3: Hierarchical Topic Dendrogram

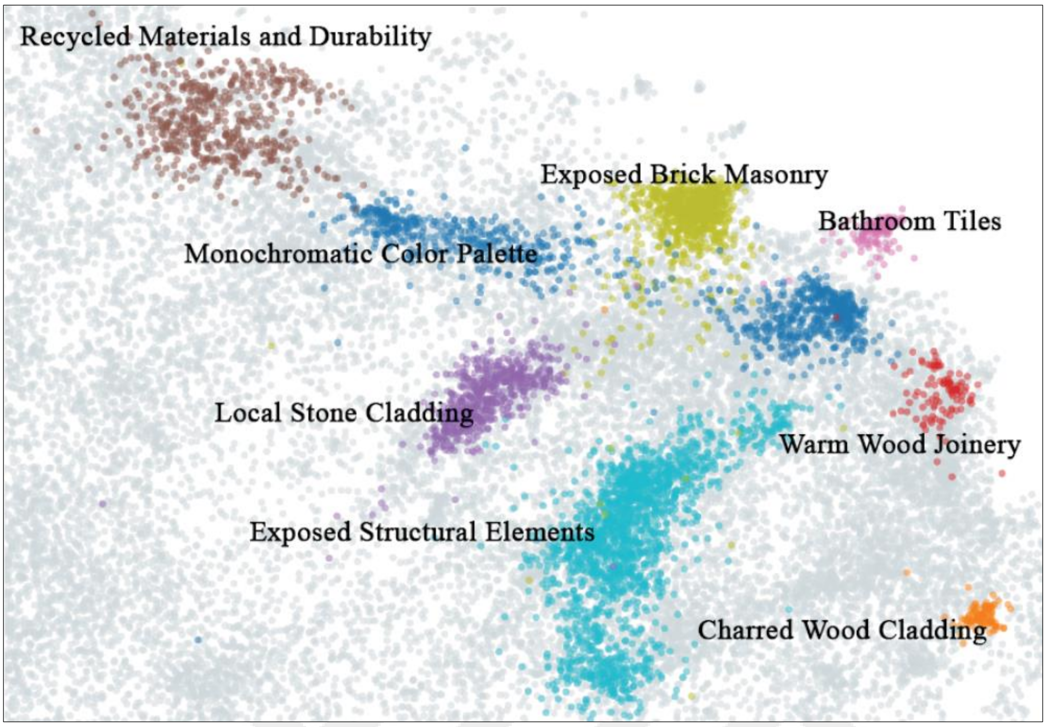


Figure 3.4: UMAP Visualization of topics under the 'Materials and Textures' aspect

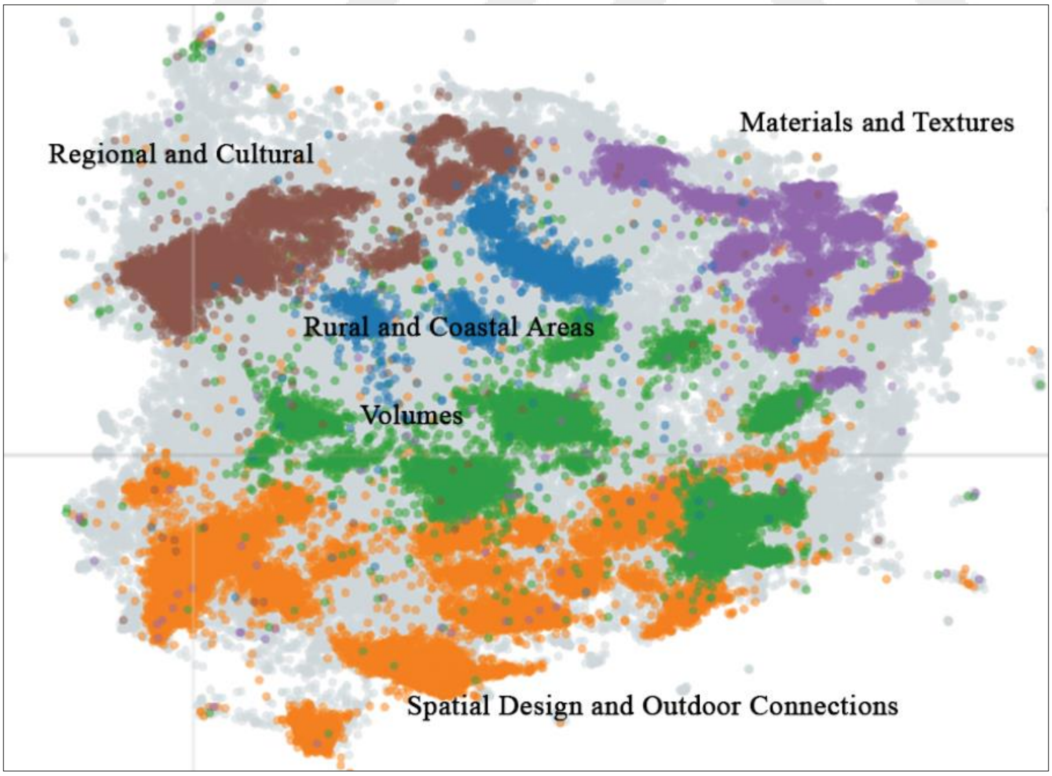


Figure 3.5 UMAP visualization of aspects

3.3.3 Creating Representative Cluster Labels

To provide a more tangible interpretation of the clusters representative labels are generated for each cluster. This process involves keyword extraction based on TF-IDF (Term Frequency-Inverse Document Frequency) metric, and generating more comprehensible labels by utilizing ChatGPT.

TF-IDF is a standard technique for quantifying word importance within a document and across a corpus. In this case, the document is the sentences that are in the same cluster. This metric assigns higher weights to words that occur frequently in a cluster (or paragraph in our case) but are infrequent across the whole corpus. Therefore, it is effectively highlighting words that characterize the specific document.

To ensure the relevance of the extracted keywords for architectural design factor analysis, certain non-informative words such as English stop words and domain-specific terms ('house', 'design', 'architecture') are excluded. Additionally, the hyperparameters of the TF-IDF vectorizer (such as the minimum and maximum word frequencies to be included in the analysis) are fine-tuned until the resulting keyword list only includes words that contribute to identifying the theme.

Table 3.1 Example topic labels generated by GPT3 based on the keyword list

Labels	Keywords
Multigenerational Living	couple, young, children, wanted, clients, live, families, generations, parents, build, son, members, needs, friends
Architectural Meaning	architectural, meaning, society, people, human, context, architects, think, timeless, object, work, conflicts
Volumetric Composition	volumes, rectangular, composition, shape, terrain, vertical, horizontal, heights, overlapping, distinct, topography
Privacy and Social Interaction	privacy, public, security, street, neighbors, residents, openness, closed, maximum, inhabitants, social, degree

Mixed Material Palette	materials, stone, wood, steel, glass, flooring, cement, exposed, tiles, including, aluminum, stones, plaster, iron
Residential Program	master, bedroom, bathroom, bedrooms, suite, children, upper, kitchen, rooms, shower, guest, dressing, dining

Given the high dimensionality of the data and the number of topics identified, interpreting the raw keyword list can be quite challenging. Therefore, the text-davinci-003 language model provided by OpenAI is employed to make the labels more comprehensible and informative. This advanced language model extends the capabilities of GPT-3, offering a greater handle on complex intents, and handling longer inputs and outputs. The model operates within the OpenAI playground, an interactive platform that offers options to save and load prompts, as well as customize model parameters to influence the level of creativity and randomness in the responses. For the label generation task, a low degree of randomness is set to ensure the reproducibility and consistency of labels. The following prompt is used for instructing the model to create representative labels:

"Below keyword sets are extracted from text clusters of architectural project descriptions of 10,000 houses on the Archdaily.com website. These are the top 50 words sorted by their c-TF-IDF scores and they represent the unique architectural features of each cluster. I want you to create a representative label for each keyword set that clearly describes the shared architectural features of that cluster. Consider the whole keyword sets holistically and come up with informative labels that can distinguish the buildings in each cluster by the architectural feature. Do not mention the non-informative concepts that are shared by all of the clusters, like "houses" or "residential buildings", "real estate", etc; only mention what makes that group of projects special and what brings them together."

3.4 Image Clustering and Contextual Analysis of Visual Features

As an additional layer of topic modeling, visual feature clusters are also utilized to enrich the architectural understanding of the generative model. The visual features of the architectural photographs are extracted by a modified VGG16 model, and extracted features are clustered with the k-means clustering algorithm. Finally, a modified BERTopic model is used for the

For the feature extraction, the final fully connected layer of the pre-trained neural network is removed. This final layer is traditionally responsible for classification tasks in the VGG16 architecture. Removing this layer leaves behind the intermediate convolutional layers that captured the abstract, high-level features from images. Next, the pre-processed images are fed into the modified VGG16 model, and features are extracted from one or more of the intermediate layers. This process outputs a feature vector for each image. These vectors are essentially a numerical representation of the visual patterns recognized by the model. With each image transformed into a 4096-dimensional feature vector, a more detailed understanding of the images' visual content is attainable.

The extracted feature vectors are clustered by using k-means clustering algorithm. This method partitions the images into k distinct clusters, where each image belongs to the cluster with the nearest mean value. The central concept of k-means is to minimize the intra-cluster variance, which is defined as the sum of the squared distances between each data point and the centroid of its assigned cluster.

To determine the number of clusters, a commonly used heuristic, 'elbow method', is employed. This method involves plotting the sum of the squared distances as a function of the number of clusters, and picking the 'elbow' of the curve as the number of clusters to use. In this study, the elbow is situated at eight clusters, indicating a point where the sum of squared distances begins to decrease at a much slower rate.

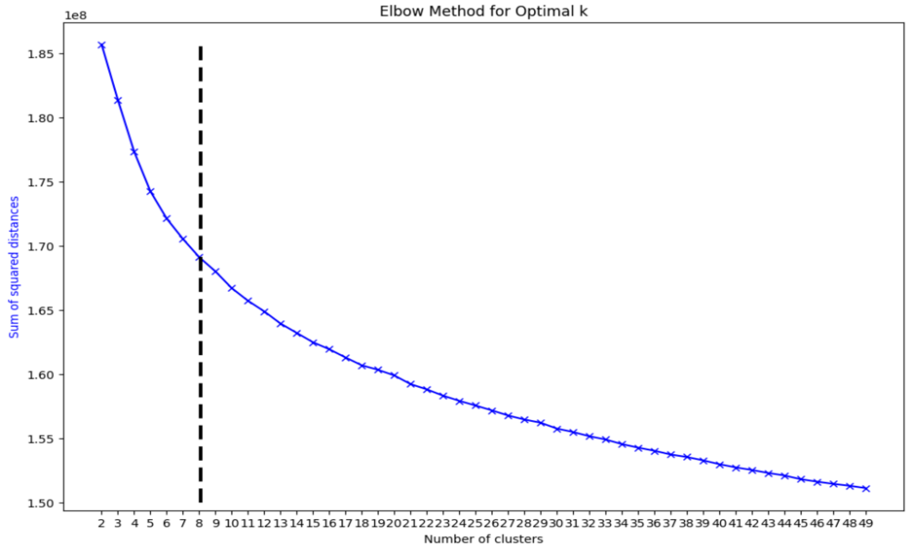


Figure 3.6: Elbow method to find the optimal cluster number based on the plot of sum of squared distances per cluster number

To extract meaningful labels for the image clusters, a modified BERTtopic model is utilized to extract keywords from sentences associated with each visual cluster. In the architectural design factor analysis, clusters of sentence embeddings were used for grouping the sentences, whereas in this case, the grouping is based on the visual clusters. The TF-IDF keyword extraction method is then applied after passing through the embedding and clustering phases. The sets of keywords obtained are used to create representative labels using the OpenAI playground.

To extract meaningful labels for the image clusters, a modified BERTtopic model is utilized to extract keywords from sentences associated with each visual cluster. In the architectural design factor analysis, clusters of sentence embeddings were used for grouping the sentences, whereas in this case, the grouping is based on the visual clusters. The TF-IDF keyword extraction method is then applied after passing through the embedding and clustering phases. The sets of keywords obtained are used to create representative labels using the OpenAI playground.

Table 3.2 Example image cluster labels generated by GPT3 based on the keywords extracted from corresponding sentences by supervised topic modeling method

Labels	Keywords
Modern Minimalist Interiors with Garden View	room, light, kitchen, interior, white, garden, windows
Urban Garden Architecture with Street Façade	site, street, city, garden, light, plot, facade, lot, urban
Living Spaces with Concrete and Wood	living, spaces, light, concrete, wood, glass, materials
Coastal Houses with Panoramic Views	landscape, located, views, natural, view, sea, land
Light-filled Open Spaces with Wood and Glass	light, wood, interior, natural, glass, large, open, garden
Forest Houses with Simple Forms and Natural Materials	landscape, views, forest, natural, trees, land, wood
Landscape-Integrated Houses with Open Interiors and Private Courtyard	garden, natural, interior, light, open, materials, level

Finally, the validity of visual clusters is evaluated through a random, uncurated selection of images from each cluster. While it is inherently challenging to objectively assess the accuracy of such unsupervised tasks, these image collages provide an intuitive gauge of the quality of the clusters. Initial assessments suggest a high degree of accuracy, as illustrated in the examples below.

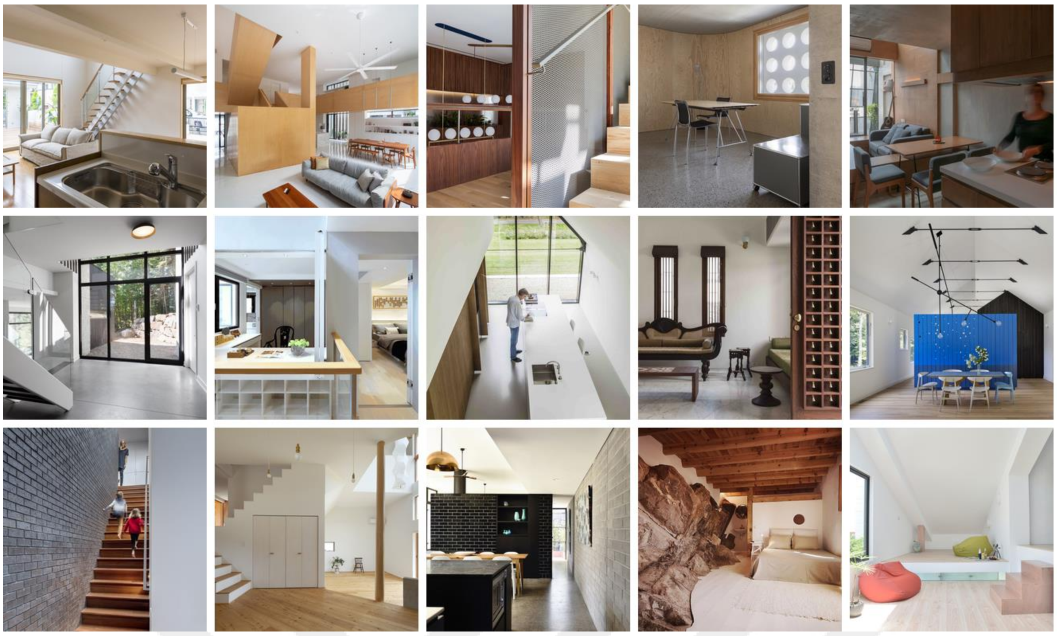


Figure 3.7 Random images from ‘Modern Minimalist Interiors with Garden View’ cluster

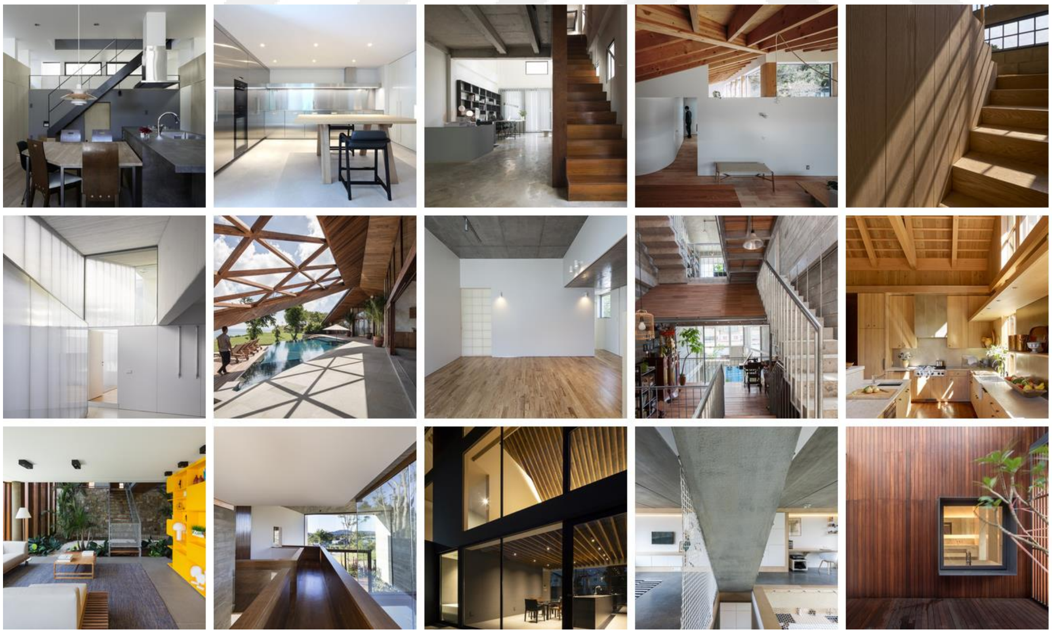


Figure 3.8 Random images from ‘Living Spaces with Concrete and Wood’ cluster

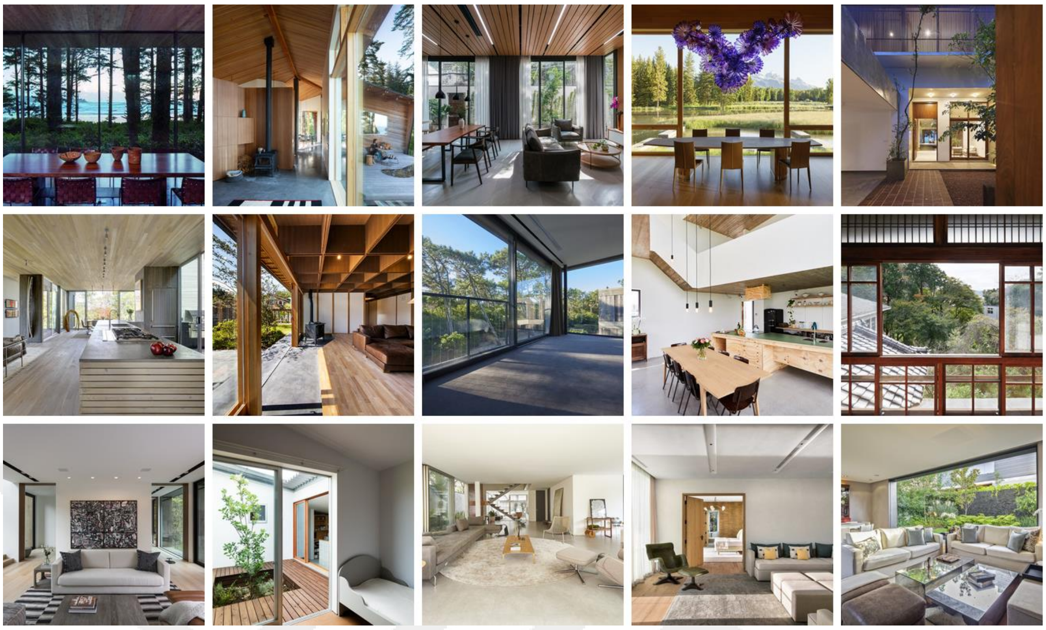


Figure 3.9 Random images from 'Light-filled Open Spaces with Wood and Glass' cluster

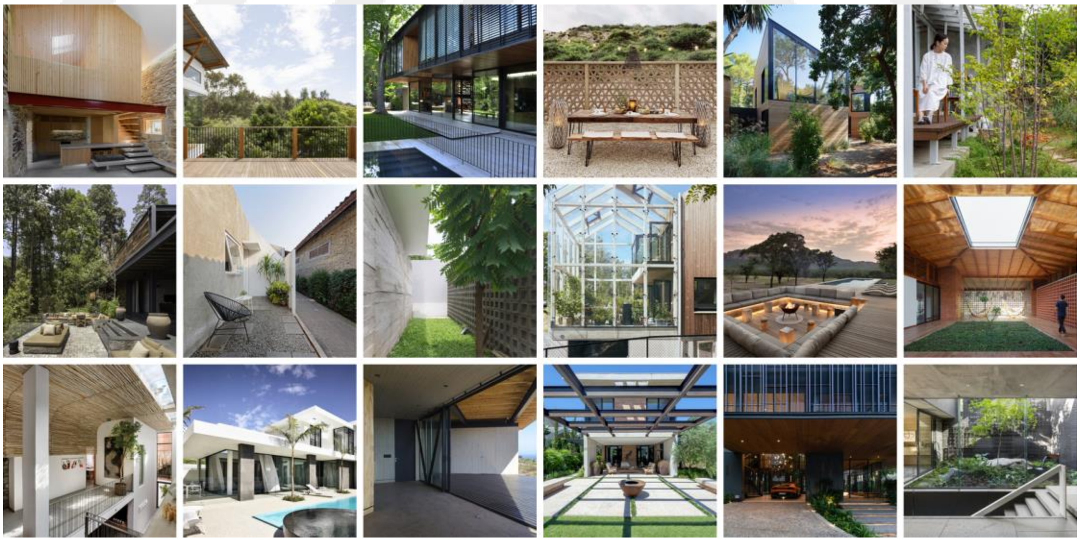


Figure 3.10 Random images from 'Landscape-Integrated Houses with Open Interiors and Private Courtyard' cluster

Figure 3.11 Random images from ‘Urban Garden Architecture with Street Facade’ cluster

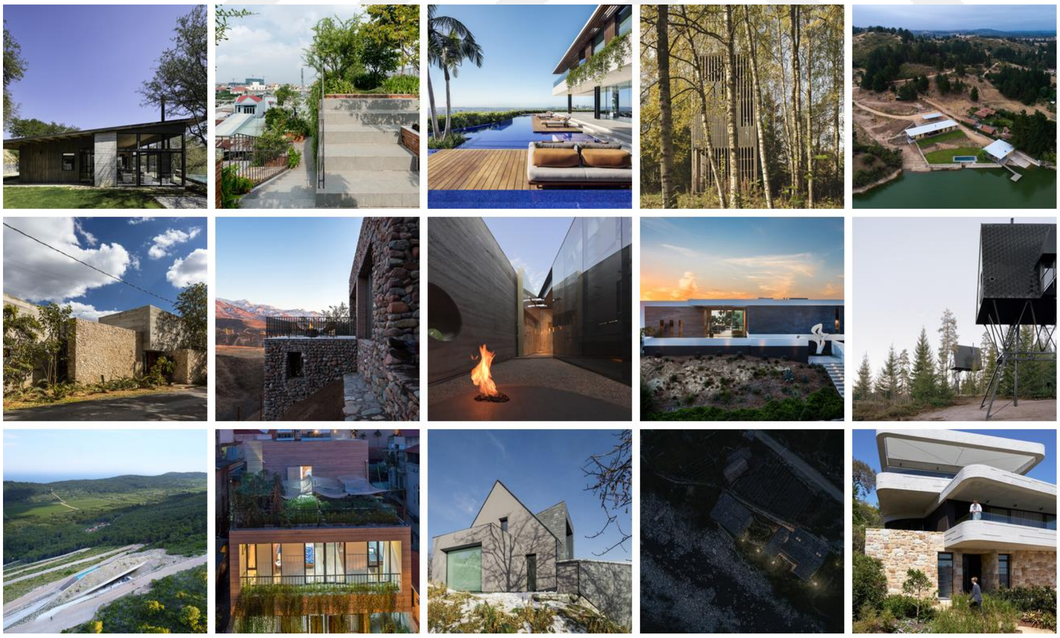


Figure 3.12 Random images from 'Coastal Houses with Panoramic Views' cluster

Figure 3.13 Random images from 'Figure 11 Forest Houses with Simple Forms and Natural Materials' cluster

3.5 Training the text-to-image Latent Diffusion Model

The Low Rank Adaptation method is utilized for fine-tuning the pre-trained Stable Diffusion model and creating a task-specific LORA module. Outputs of the design factor analysis (3.3) are used as labels for the training data in order to instill a structured understanding within the latent diffusion model, which is traditionally trained on billions of unstructured image-caption pairs.

This approach leverages the extensive visual knowledge inherent in the stable diffusion model. At the same time, it guides the model to learn specific architectural design factors extracted from a carefully curated and traceable dataset. This dual strategy aims to enhance the model's comprehension of architectural design and improve its interpretability.

The training data consist of images and corresponding captions which are structured in a specific format to guide the model in learning visual features that correlate with the extracted architectural design factors. When this format is used to prompt the model, it generates architectural images that reflect these design factors, resulting in outputs that are both nuanced and contextually aware.

3.5.1 Preparation of Image Caption Pairs

To prepare the structured captions, various labels derived from topic modeling, image clustering, and project-level information are transformed into image-level information by associating sentences with images based on their relative positions on the project web page.

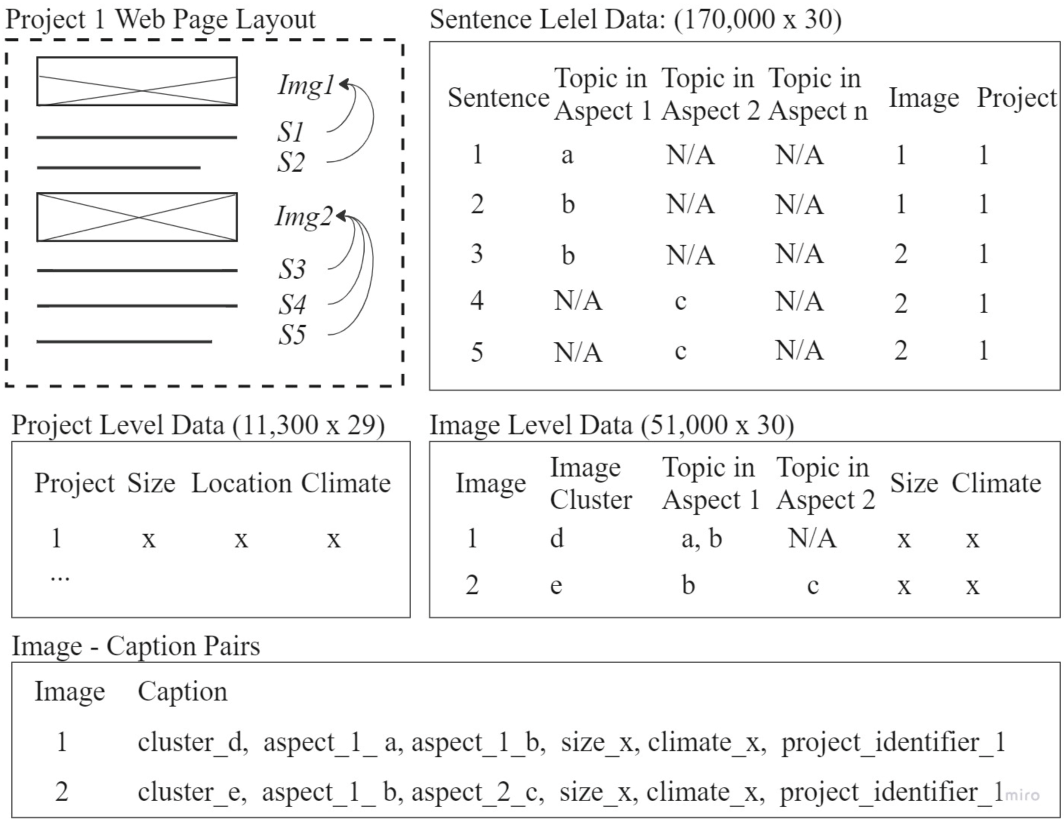


Figure 3.14 Diagram showing the preparation of image-caption pairs

The structure of the captions follows a specific format. Firstly, they include the applicable design factors of an image, delimited with commas. The total of 21 aspects encompasses the extracted topics within the 15 aspects, project size, visual clusters, region, and the discretized geographical features such as average temperature, and rainfall.

Additionally, a project identifier is included to account for situations where multiple images within a project focus on specific design factors. This allows the model to learn and consider the factors of an image based on the collective information from other images within the same project. It is also useful for evaluating how well the model can generate similar images to the original data.

During the training process, each factor in the caption is shuffled to prevent the model from learning any specific order that lacks meaning. Furthermore, the probability of the assigned topic to the corresponding sentence of the image is included in the caption. This addition enables testing the visual correspondence of the assigned topic.

An example prompt of a structured caption is as follows:
size_4,city_size_5,wealth_1,rain_4,temp_5,reg_south_asia,probs_0.3,image_type_6_Landscape_Integrated_Houses_with_Open_Interiors_and_Private_Courtyard,aspect_11_volume_Connected_Void_Spaces,aspect_14_construction_Prefabricated_and_Modular,nl_prj_identifier_10

3.5.2 Fine-tuning Stable Diffusion with Low Rank Adaptation Method

Fine-tuning is a widely employed technique in natural language processing and computer vision domains to adapt a pre-trained model to a specific task or domain. It entails taking a pre-trained model, previously trained on a large-scale dataset, and further training it using a smaller dataset specifically curated to the target task.

The process of fine-tuning large language models that contain billions of parameters is both computationally expensive and challenging due to a phenomenon known as catastrophic forgetting. However, the Low Rank Adaptation (LoRA) method provides a solution to these issues. It enables efficient training of these large models within limited memory resources, while also solving the catastrophic forgetting.

The LoRA method leverages on the low-rank structure in the parameter space of neural networks. Rather than fully fine-tuning the base model's weights, it freezes the pre-trained layers and adds a select number of trainable rank-decomposition matrices, known as LoRA modules (Hu et al., n.d.). These compact, task-specific LoRA modules allow easier transitions between tasks. In this study, the LoRA checkpoint's weights amount to a mere 36 MB, much smaller than the 4 GB base model Stable-Diffusion-v1-5.

CHAPTER 4

RESULTS AND EVALUATIONS

The performance of the trained model is evaluated across four key dimensions. The initial phase scrutinizes the Contextual Relevance of Extracted Architectural Design Factors, which establishes the foundation for subsequent evaluations. Here, the accuracy of the hierarchically clustered architectural topics is evaluated through semantic relations and the concurrence of image-sentence pairings.

Next, the Task-specific Adaptation Performance is evaluated, which ensures the effectiveness of the fine-tuning process. This stage assesses the model's capability to generate images based on the introduced domain knowledge and distinguishing itself from the base stable diffusion model. The performance is validated through image comparisons from the fine-tuned and base models, using identical prompts.

The third evaluation delves into the correlation between generated outputs and their corresponding training images to understand the extent to which features in the input images influence the generated images. This is carried out by comparing generated visuals with original images refined using the same parameters. Moreover, this evaluation endeavors to shed light on how the model's response varies based on the specificity of the prompt, exploring the nuances between design research and unique design inception within the realm of generative AI.

Finally, the relevance of the generated images investigated, aiming to ensure that the model correctly interprets and captures the architectural intent conveyed in the prompts.

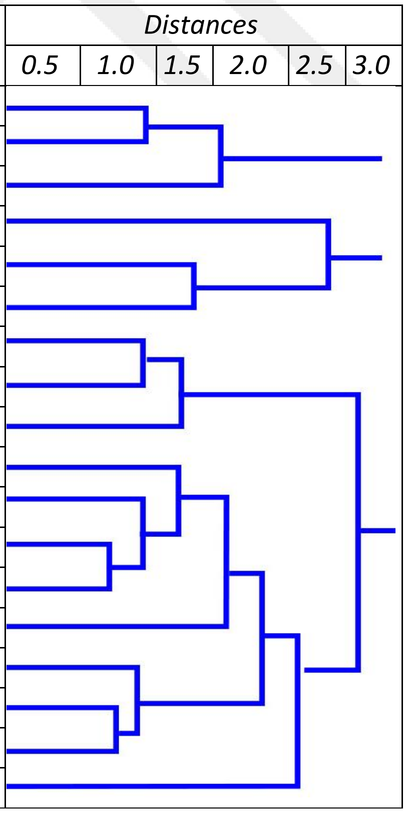
4.1 Coherence of the Extracted Label Structure

The hierarchical clustering of architectural project description sentences has resulted in 114 distinct topics under 15 overarching design aspects. The multi-tiered structure of the design factors is presented through tabular representations and hierarchical clustering visualizations. The system-level accuracy of the structure is discussed through observed semantic relations. The topic level accuracy is discussed through uncurated collages of image & sentence pairs of topics.

Upon initial analysis, the organization of the extracted topics suggests a promising degree of accuracy within the model. In the primary tier, topics are predominantly classified into three key categories: 'Design Process', 'Building Features', and 'Cultural and Environmental Influences'. These categories contain multiple subtopics, some of which delve deeper into more precise areas of interest.

Table 4.1 Hierarchical topics grouped by main design aspects

Labels generated by OpenAI text-DaVinci-003		
Design Process	People	Design Process and Requirements
		Meaning and Experience
		Conceptual Design
Building Features		Materials and Aesthetic Choices
	Spatial Planning	Functional Areas
		Spatial Composition Transition
Cultural and Environmental Influences	Regional and Cultural Context	Tropical climate
		Urban Development
		Rural and Coastal areas
	Landscape Interactions and Design Elements	Façade Cladding
		Nature Integration and
		Views and Outdoor Spaces
		Form and Functionality
		Service Spaces
	Influences	Rural Influences
		Artistic Influences
		Manufacturing Methods
		Sustainable Strategies and Energy Efficiency



The first category primarily revolves around the intangible aspects of the design process and focuses on the conceptualization and thought process involved in the creation of an architectural project. The topics in this group explain how the design idea was conceived, what kind of challenges were faced, and how they were overcome. It delves into the architects' thought process, the main design questions, and the main design considerations. It also includes the overall philosophy or intention behind the design, such as the focus on family, privacy, or harmony.

The 'Building Features' category focuses on the tangible, physical elements of architectural design. It details the physical manifestation of the design process, the materials used in the construction, and the resulting architectural features of the building. This group is very grounded and deals with concrete, observable aspects of the building. It covers a wide range of elements from the specific type of roofing or cladding used, to the fixtures in the bathroom, or the arrangement of the interior spaces.

The third category of topics, 'Cultural and Environmental Influences', looks at architectural design from a broader perspective, taking into account the context of the building both in terms of cultural/regional influences and environmental considerations. It's a zoomed-out view of the design process, focusing on how the design interacts with its surroundings and the broader societal and geographical context. It starts at a macro scale by considering global or regional architectural styles and trends, then narrows down to the specific site or plot, and how the building interacts with its immediate environment. This group also includes considerations for sustainability and energy efficiency.

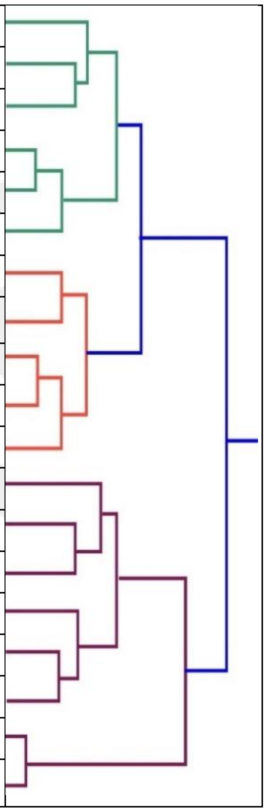
It is important to note that the interpretation of this design factor system retains a degree of subjectivity because of the inherently creative and interpretive nature of the architectural design. Nevertheless, by delivering these factors in a coherent and accessible format, the intention is to demonstrate the capabilities of neural networks to capture design factors and provide users of generative models with a deeper understanding of how specific design factors can influence the algorithm.

4.1.1 Design Process

The first group of topics navigates between two central axes: 'People', which encompasses the social, perceptual, and experiential aspects; and 'Conceptual Design', which captures the main technical, formal, and environmental decisions.

Table 4.2 Hierarchical topics related to the design process

Design Process (52,000 Sentences)	People	Design Process & Requirements	Architectural regulations and exceptions
			Building history and renovation
			Family and Multigenerational Living
			Architectural Meaning
			Problem-Solving Design
			Architect-Client Collaboration
		Meaning and Experience	Human Condition and Existential Reflection
			Harmony of Humans and their Surroundings
			Human-Centric Landscape Integration
			Spatial Perception
			Spatial Sequence and Experience
	Conceptual Design		Native Vegetation-Rich Landscapes
			Connection and Privacy of Gardens
			Sunlight and Shadow Play
			Geometric Facades
			Sloped Roofs Geometry
			Volumetric Composition
			Visual Barriers for Privacy and Security
			Privacy and Social Interaction



The 'People' axis reflects the human-centric aspects of architecture. Themes that are ncapsulated within ‘Design Process & Requirements’ are serving as the backbone of architectural intent and outline the motivations and driving forces behind a project.

Upon examination at the topic level, the structure exhibits coherence and logic. For instance, while the 'Architectural Meaning' and 'Problem-Solving Design' themes represent distinct design motivations and are closely linked to the model, the 'Architect-Client Collaboration' theme is connected to both, serving as the mutual platform where these two design motivations emerge. Augmenting the 'People' axis, themes associated with 'Meaning and Experience', such as 'Human Condition and Existential Reflection' and 'Spatial Perception', delve into socio-cultural considerations in architecture. While these themes still represent high-level intentions or focuses, they provide an in-depth exploration of the human dimension within the architectural discourse.

Table 4.3 Random image-caption pairs: 'Design Process' / 'People' / 'Meaning and Experience'

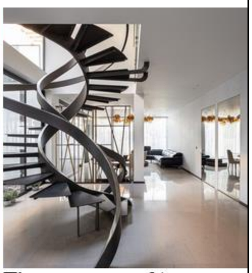
Spatial Perception			
			
A subtle and constant link between exterior and interior spaces, blurs the perception of spatial borders.	2.The exposed concrete finish of the house adding to the ethnic theme of the residence.	A multi-dimensional place is created in this continuous space.	As on the exterior, the interior is also designed with the thought of its relation to the surrounding nature.
Human Condition and Existential Reflection			
			
For Patrick and Hugues, Origins was an opportunity to create a global concept from scratch, down to its ultimate detail	A journey... to lose the identity and discover self.	Roof rafters are exposed on ceiling to describe traditional roof structure of stone house in Jeju in the same way.	The canopy of trees from the green belt gives an illusion of an infinite green to the terrace garden.

Table 4.4 Random image-caption pairs from the topics in ‘Design Process’ / ‘People’ / ‘Process and Requirements’

Problem-Solving Design			
			
It takes advantage of the lay of the land and the movement of the sun and contributes to the formal rhythm and spatial condition of the street.	The volumes and voids formed as a resultant of the superimposition of the tilting grids, are marked by the prominence of broad brick walls.	These were the key aspects which informed our response.	Duringthe first visit to the place, the team already realized that this project would be special.
Architectural Meaning			
			
The experience of the " Khaneye Hayatdar " for us, was an experience of redefining the soul of Iranian architecture with an avant-garde appearance.	A simple, careful, minimal, essential construction can lead us to a good architecture.	Inspired by the writings of the transcendentalist Emerson and Thoreau, the cabin stands as an example of introverted architecture or neo-transcendentalism.	There is a wide variety of apparent materials that will age with dignity over time and will blend with the surroundings.

Concurrently, the 'Conceptual Design' axis explores the architectural concepts that crystallize these abstract ideas into tangible form. Themes like ‘Geometric Facades’, ‘Sloped Roofs Geometry’, and ‘Volumetric Composition’ speak to the formal articulation of architectural intent, dealing with the aesthetics and physical shape of buildings. Similarly, themes like ‘Native Vegetation-Rich Landscapes’, ‘Connected and Private Gardens’, and ‘Sunlight and Shadow Play’ are particularly, demonstrate how a building's design engages with its environment.

Table 4.5 Random image-caption pairs from the topics in 'Conceptual Design' category

Volumetric Composition:

			
My design has the intention to amplify the character of each area.	the south facade in the direction of the slope then illustrates the volumetry in its entire two-story expression.	The two housing volumes rise in order to dominate the surroundings from the interior.	Detached from the sides by a glass floor, the volume gains a floating aspect in the double-height and brings variance in the scale of the room.

Privacy and Connections of Gardens

			
The sunken yardÃ¢â,¬Â“ usually meant to allow light into a basement floor and mostly placed at the fringes of the plotÃ¢â,¬Â“ is this house's central motif.	The entire house opens to the outside, which provides a cross-ventilation flow in the environments, and the big mass of vegetation collaborates for a great thermal comfort	Both the continuity of the park and the artificial top are poetically intertwined in a deep spatial synthesis.	The relatively constrained site sits just off the road between Reading and Oxford in West Berkshire, midway between the two hospitals where both

Native Vegetation Rich Landscapes

			
There, a bosque of aspen trees provides shadow and texture and acts as a focal point within the natural clearing.	Small orchards will take advantage of the rains during and after the rainy season.	The west side of the island topography is characterized by irregular sharpen slopes dominating the Adriatic see.	Some plants were relocated as if no work had been done in its 10-month duration.

4.1.2 Building Features

The ‘Building Features’ group branches into three discernable clusters: 'Materials and Aesthetic Choices', 'Functional Areas', and 'Spatial Composition and Transition Elements'. Together, these clusters provide a more granular examination of the building's physical and spatial characteristics. In its structure, it is more visually and materially oriented and its focus is narrowed down to the building itself without mentioning any environmental or conceptual consideration.

Table 4.6 Hierarchical topics related to building features

Building Features (55,000)	Materials and Aesthetic Choices		Steel and Corrugated Metal Roofs
			Exposed Structural Elements
			Exposed Brick Masonry
			Local Stone Cladding
			Sustainable Timber Cladding
			Mixed Material Palette
			Recycled Materials and Durability
			Bathroom Tiles and Fixtures
			Monochromatic Color Palette
			Warm Wood Joinery
	Layout	Program	Residential Program
			Flexible Program
			Kitchen as Socializing Area
		Spatial Design and Outdoor Connections	Height Perception
			Skylight Illumination
			Circulation and Spatial Connections
			Flexible Arrangements
			Framed Window Openings
			Sliding Doors
			Swimming Pool
			Deck-Extended Outdoor Living
			Terraces and Balconies
			Inner Courtyard-Ventilated Rooms
			Inner Patio-Circulation Generating

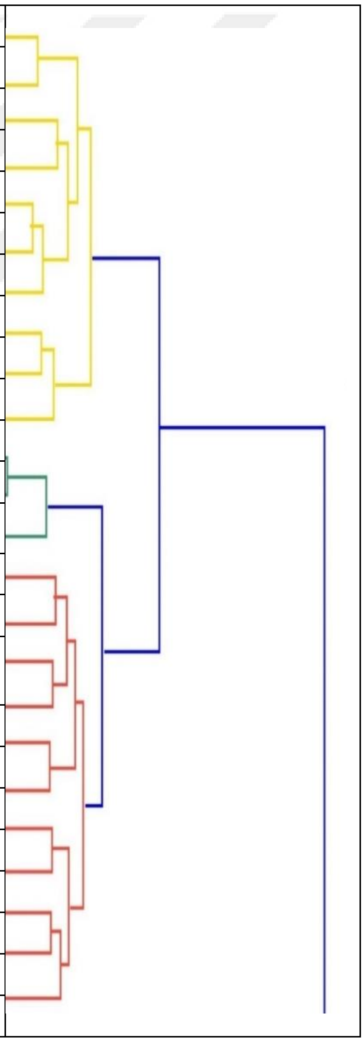


Table 4.7 Random image-caption pairs from the topics in 'Materials and Aesthetic Choices' category

Exposed Structural Elements			
			
The definition of the object does not require detail or termination, it is intrinsically the structure.	Concrete is a very moving material.	In Bunker House, we employed this concrete system not only as structure, but also as spatial partitions and	A series of orthogonal timber forms sit inside the concrete shell, easing the more functional requirements
Local Stone Cladding			
			
Respect the stone, recover the existing elements and combine them with an open and new distribution, actual lighting and furniture, creates a new charming	The overall feeling of the three-storey residence in Melbourne, Australia is lightness - almost an ethereal floating quality created by the sun refracting over the	A granite basement that is embedded in the ground, a stone pedestal that in the front blocks the facade but that in the back makes it disappear and becomes lighter: a beam	The initial idea is to allow habitant's behavior to carve a dwelling space into a monolithic piece of marble sculpture.
Mixed Material Palette			
			
The texture of high-quality materials provides a richness that greatly enhances the attraction of the space.	Besides creating small congregational areas, this staggering also gave the design team opportunities to expand the amount of garden space within the internal spaces.	The greenery makes the upward journey a pleasant one, as the overlapping plants invite you to touch them, creating a pleasurable sensation of walking in a garden.	The materials are from the region: brick, wood and earth.The earth is from the site, the one that was dug to bury the foundations was refused as finished on all the walls.Thus, architecture emerges

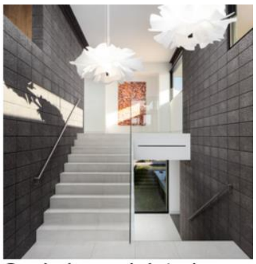
The 'Materials and Aesthetic Choices' cluster emphasizes the visible and tactile qualities of architecture. The topics in this category are organized predominantly by visual and physical correlations.

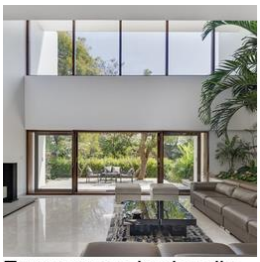
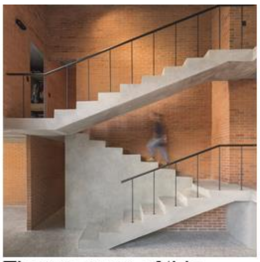
In comparison, the 'Spatial Composition and Transition Elements' category are organized into three groups that show different levels of visual correlations. General topics such as 'Height Perception', 'Skylight Illumination', 'Vertical Circulation and Spatial Connections', and 'Flexible Arrangements encapsulate broad design aspects without articulating specific characteristics. Consequently, an immediate visual correlation within their corresponding images is not apparent, as illustrated in. However, topics referring to specific architectural elements, such as 'Framed Window Openings', 'Sliding Doors for Seamless Connection', and spatial architectural design elements like 'Decks', 'Terraces', 'Courtyards' are visibly coherent within the collages in Figure 27, Figure 28, and Figure 29.

Analysis of the logical relationships between the extracted topics demonstrates the underlying rationale within the clusters, that aligns with architectural considerations. Topics such as 'Height Perception', 'Skylight Illumination', 'Vertical Circulation and Spatial Connections', and 'Flexible Arrangements' are all related to space and movement and they are closely linked by the algorithm. Similarly, topics like 'Swimming Pool-Centered Entertainment', 'Deck-Extended Outdoor Living', Terraces and Balconies', 'Inner Courtyards', and 'Inner Patios' are all related to particular spatial architectural elements, and they are closely linked.

Table 4.8 Random image-caption pairs from the topics in ‘Spatial Composition and Transition Elements’ category

Skylight Illumination			
			
We see the beauty of this and want to enhance this "windowshuffle" even more by adding skylights.	Contrary to what one might think, this semi-buried house was brighter and more sunny than a conventional house.	The landscape ofÃ…â€œThe Pucuk Mer treesÃ‚Â□ functions as barrier sunlight, and creatingÃ‚Â a micro-climate in the periphery of the site providing micro-	the practically complete demolition and reconstruction of the existing dwellingÃ¢â‚¬Â□ a large four-storey end-of-terrace house,

Height Perception			
			
The current structure includes an entre-sol with a low ceiling, and in the new design we favor the differences of floor cos (height), so changing the height of the entre-sol is necessary.	Preserve your house, or make a new one with a high long term value, and you make your house more environmentally friendly.	The white volume, with greater entity, and two floors, houses the double height access, the Living room, dining room, kitchen-laundry area, on the ground floor, and on the first floor is the main room	Scale is modulated between cozy and dramatic via the inclusion of floating ceiling clouds and light coves that define the spaces below.

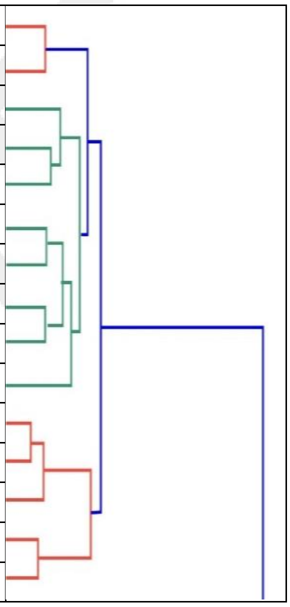
Flexible Arrangements			
			
Every space is visually connected to enhance interaction within the house.	A house of unified divisions.	The program of this single-family dwelling is resolved in four staggered volumes each with a 5 x 5 meter footprint.	Work to House 108 involved carefully stitching new spaces into existing spaces.

4.1.3 Cultural and Environmental Influences

The third topic group approaches the architectural design process from a more holistic viewpoint. It captures the influence of larger contextual factors including cultural, regional, and environmental aspects on design decisions. This perspective starts at the macro level with global or regional architectural styles and trends. It then transitions to a micro level, focusing on how the particular building interacts with its immediate surroundings. Within this category, sustainability emerges as a significant component.

Table 4.9 Hierarchical topics related to cultural context

Cultural, Environmental, and Sustainable Architectural Influences (60,000)	Regional and Cultural Context	Tropical climate	Climate-Adaptive Architecture
			Tropical Modernism
		Urban Development	Urban Boundary-Defined Plots
			Plot Size and Dimensions
			High-Density Urban Suburb
			Vietnamese Urbanization
			Thai Suburban Lifestyle
			Japanese Residential Renovation
			Chinese Courtyard Residences
			Korean Townhouses
		Rural and Coastal areas	Rapid Urbanization of Villages
			Nordic Suburban Development
			Latin American Coastal Residences
			Panoramic Mountain View
			Beachside Relaxation



Within the ‘Regional and Cultural Context ' cluster, a clear division is apparent between urban and rural areas. Upon initial examination of Tables 4.8 and 4.9, it becomes evident that topics within this category category exhibit a certain level of visual correlation. However, this correlation is different from the ones observed within the 'Building Features' cluster since the shared characteristics in this cluster are not as directly apparent or visually pronounced. This highlights the more nuanced and diverse influences that culture and region exert on architectural design.

Table 4.10 Random image-caption pairs from the topics in 'Urban Development' category

Climate-Adaptive Architecture			
			
Located in the upcountry dry zone region of Sri Lanka about 2 kilometers away from the Diyathalawa military training school, is a cozy abode designed by Architect Damith Premathilake.	The region is arid, characterized by mild winds blowing from the coast inwards.	It is a must-have in a crowded tropical city, with extreme temperatures, heavy traffic and alarming air and noise pollution.	TheÃ…â€œEsquina VerdeÃ‚Â□ project was conceived as a house that could eventually evolve into a rental lodge for tourists and explorers, who constantly visit the area because of it`s tropical rain forests
High-Density Urban Suburb			
			
Lines are the basis of any structure and to keep the purity of form intact we decided to use a play of lines for the design, Starting in two dimension and then continuing to use it to create a strong visual	It's an upper-class country dwelling.	Located in Bandung, Indonesia, The FE House is a serene home where every element is designed around family and the location's tropical climate.The house consists of spacious living space	Living in a tropical climate like Indonesia brings a lot of perks for residential design, especially the high sun exposure and rainfall each year.
Urban Boundary-Defined Plots			
			
TheÃ‹Å“L' shaped site measuring 3038 sft is flanked by roads on two sides.	The site is located in the local city developing gradually.It faces the 30-meter-wide main road, increasing	This residence is an ideal illustrative of an urban home designed for a family of four.	the plot's location is serene and descends towards the east-northeast.

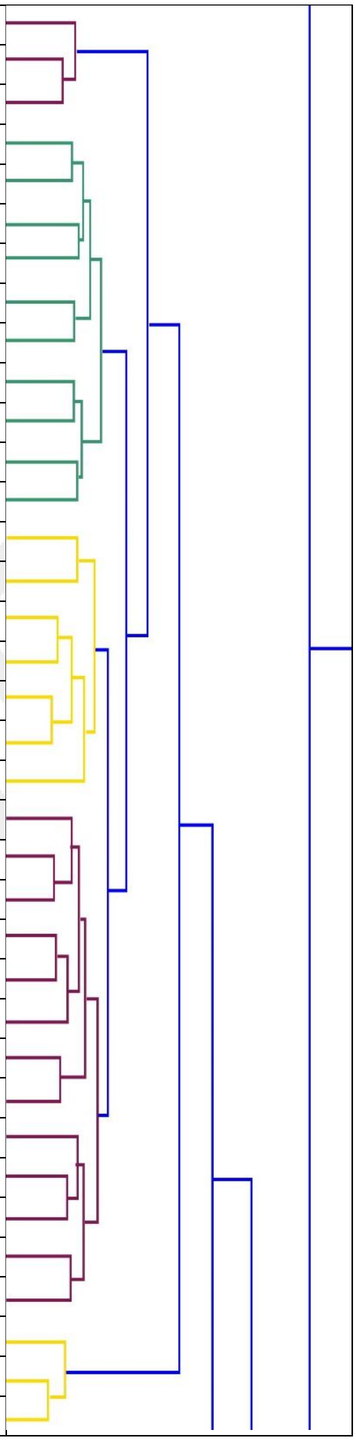
Table 4.11 Random image-caption pairs from the topics in 'Rural and Coastal Areas' category

Latin American Coastal Residences			
			
We find this house with simple linesjoining the catalogue of architectural and almostsculpturalstyles that surround it, located in a private urbanizationofan American spirit on the outskirts of Zaragoza.	Located on the outskirts of the city of Cancun, Mexico, the architectural program distributes the space into three levels, including: living room, dining space, kitchen, backyard patio, storage room, four bedrooms,	Located near the Tigre Delta, the house has the peculiarity of being situated in a floodable area, with internal channels and a forest that mostly invades the whole, but being on the side of the continent.	A typical beach house, house 77 is situated on a condominium site on the north coast of the state of Rio Grande do Sul, situated 130km from Porto Alegre.
Panoramic Mountain View			
			
The Aculco project is a holiday home completely disconnected from the city in the middle of nature.	Unexpectedly, this calculated geometry relates almost candidly to the nearby hills topography.	The Tai Tapu house is situated on a west facing hill on the lower slopes of the port hills, about 5kms out of Tai Tapu.	Concept:The land is situated in Cerdanya's area, in an old settlement where rehabilitation began 15 years ago.

The 'Landscape Interactions and Design Elements' cluster focuses on the relationship between the building and its immediate natural surroundings. At the heart of this interaction, the cluster labeled 'Nature Integration and Sustainability explores the engagement of the building with the site's unique attributes, as demonstrated by topics like 'Sloped and Embedded Topography', ‘Wilderness Camping Cabin’ and ‘Fencing and Boundary Engagement’

Table 4.12 Hierarchical topics related to landscape interaction and design elements

Cultural, Environmental, and Sustainable Architectural Influences (60,000)	Landscape Interactions and Design Elements	Façade Material	Charred Wood Cladding
			Stained Wood Siding
			Patinated Metal Cladding
		Nature Integration and Sustainability	Bamboo Network for Privacy
			Fencing and Boundary Engagement
			Sloped and Embedded Topography
			Formal and Enclosed Courtyards
			Gabled Roofs Reinterpretation
			Rainwater Harvesting Systems
			Water for cooling and reflection
			Urban Contemplation Retreat
			Wilderness Camping Cabin
			Lakefront Orientation
		Views and Outdoor Spaces	Luxury Villa in Natural Setting
			Operable Shutters for Solar Control
			Glazing and Expansive Views
			Balcony-Protected Outdoor Living
			Pergola-Shaded Outdoor Spaces
			Hillside Cantilevered Structures
			Screened Porch-Protected Entry
		Form and Functionality	Veranda-Connected Indoor-Outdoor
			Abstract Cubic Prismatic Forms
			Floating Wooden Boxes
			Glass-Clad Pavilions
			Bridged Connections
			Winged Layout
			Fireplace as a Focal Element
			Loft-Style Bedrooms for Children
			Multi-Functional Mezzanine Design
			Atrium for Ventilation and Lighting
			Connected Void Spaces
			Circulation Axis
			Golf Course Estates
		Service Spaces	Vehicular Access
			Garage and Entrance
			Basement Amenities



The 'Views and Outdoor Spaces' cluster delves into more building-centric aspects while retaining the emphasis on the environment. It explores how the building communicates with its surroundings through visual channels, as can be seen by topics such as 'Glazing and Expansive Views' and 'Hillside Cantilevered Structures.'.

Table 4.13Random image-caption pairs from the topics in 'Views and Outdoor Spaces' category

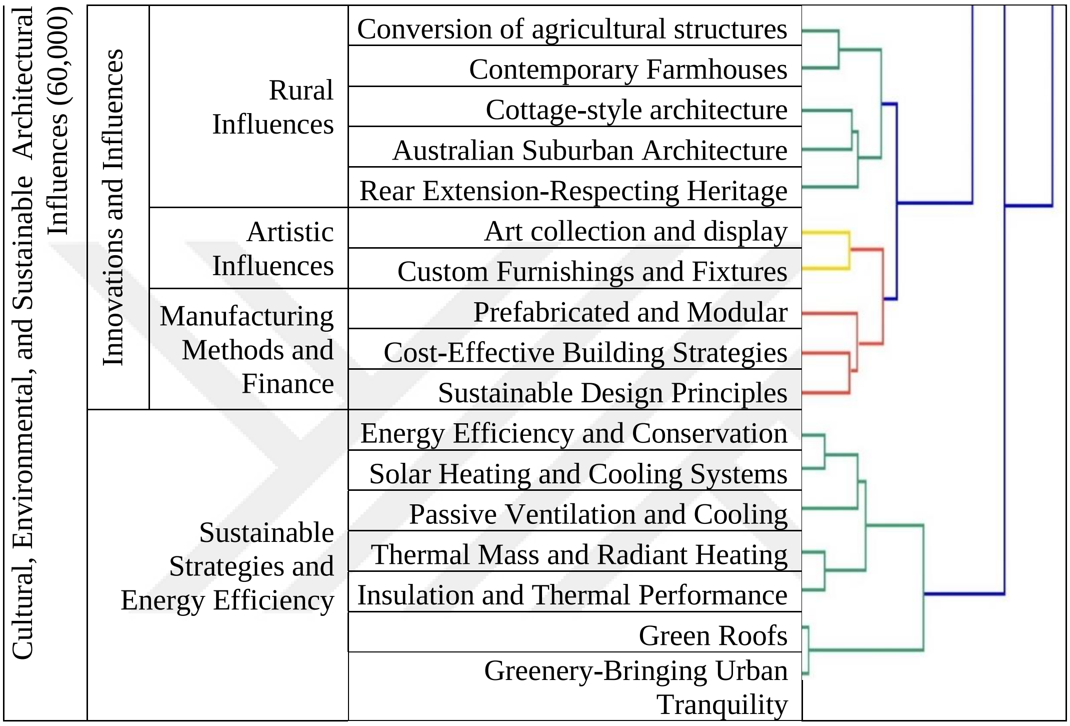
Wilderness Camping Cabin			
			
Their grown children work as natural resource scientists and spend much of their time deeply engaged with the outdoors.	A couple from the Caribbean seek retreat amid the northern Andes in a small town, two hours north of Quito, an orchard site in a valley 2470 meters	Ephemeral Edge is a weekend retreat and retirement home located three hours north of New York City.	After a very busy career Nico and Hilde wanted to create a haven of peace on this quietly located plot.
Formal and Enclosed Courtyards			
			
A small and lonelyhackberry tree on the front of the plot dictated the definite location of the house and was the organizer of the final composition.	The goal of the architect was to largely preserve the natural slope and to intervene as little as possible.	We took the slope of the land defined the movement of soil, half-buried, and that the project is somewhat introverted dressed in the texture of the site, its vegetation,	Carrying out the house included an excavation and an important earth removal, for which this natural resource was used to make blocks of compressed earth on th site.

Gradually shifting focus, the 'Volumes, Form, and Functionality' cluster navigates through the architectural structure itself, examining how it settles into the site. One axis focuses on the building's exterior volume, as represented by topics like 'Abstract Cubic Prismatic Forms' and ‘Connected Void Spaces,' another axis focuses on the internal volumes, as represented by themes such as 'Atrium for Ventilation and Lighting' and 'Multi-Functional Mezzanine Design.'

Table 4.14 Random image-caption pairs from 'Form and Functionality'

Connected Void Spaces			
			
The street-side room serves as a vestibule, a reception space plus a storage room.	The living room develops in length, parallel to the sea.	The bedrooms on the two lower floors are accessed by two opposing staircases.	Each room is open on many sides whilst limiting the number of corridor spaces.
Abstract Cubic Prismatic Forms			
			
The reflective roof of the deep veranda brings back garden view to the interior.	FH1 house was designed for temporary and permanent stay, is intended to provide both a connection to the outdoors and a	For the fully integrated spaces, the trio has chosen a shelf to delimit the space of the bedroom.	To the north of the municipality of La Tour-de-Peilz, on the edge of Lake Geneva, lies an old vineyard above the town.
Multi-functional Mezzanine Design			
			
have improved the performance of houses, but rather than simply relying on such individual functions, we would like to think about structures that work better when they	The natural sunlight from the sky lit above the staircase illuminate not only staircase but also fist floor corridor as it leads to the guests to the second floor.	It was necessary to redo the full coverage, taking advantage of the attic to the roof for adding a new bedroom floor.	By using transparent materials, this multipurpose area is connected habitants to the joy of the city.

The final clusters, 'Innovations and Influences' and 'Sustainable Strategies and Energy Efficiency', display a higher-level interaction with the 'Landscape Interactions and Design Elements' cluster. Instead of aligning within the gradual scale shift in the 'Landscape Interactions and Design Elements' cluster, these overarching categories resonate uniformly with all levels of the preceding topics.



This cross-scale correlation corroborates the validity of the hierarchical clustering method and demonstrate its efficacy in discerning reasonable connections among the topics based on their contextual relevance.

4.1.4 Summary of the Findings

In the system-level evaluation, the hierarchical multi-tiered structure of the topics demonstrates a general logical coherence, as revealed through discrete examples. It suggests that a structural and formal reasoning perspective about a particular domain can emerge from data without explicit human intervention. As outlined in the background chapter, this methodology resembles common ad-hoc explanatory methods for machine learning algorithms, which typically involve capturing an event with neural networks and applying formal reasoning or using inherently interpretable methods to derive formal knowledge. In this study, a deep learning tool captured the semantic relations, and an ad-hoc interpretation was applied.

In the topic-level evaluation, two distinct aspects emerged (1) They may represent a specific physical architectural characteristic or a broader concept that underscores the design focus rather than the final design's physical attributes. (2) A topic may exhibit high or low visual correlation, regardless of its specificity. For instance, topics with low specificity, such as ‘meaning’ or ‘problem-solving’ presented a readily observable correlation to the human eye. In contrast, topics with high specificity, such as "sliding doors" or "atriums," did not present a strong visual correlation.

These subjective observations depend on limited examples. While human capabilities for pattern recognition and abstract reasoning far surpass those of any neural network algorithm, the ability to comprehend thousands of examples simultaneously is unique to neural networks. Hence, there may be correlations or shared characteristics in the images of a particular topic that may go unnoticed by human observers but are captured by the algorithm, or vice versa.

4.2 Task-specific Adaptation Performance

The fine-tuning phase involves embedding the extracted design factors into the text encoder of the stable diffusion base model as unique concepts. This integration ensures that the fine-tuned model, if effectively trained, becomes distinct from the base model, generating images based solely on the input visuals and the structured design criteria outlined in this study.

To evaluate the success of the fine-tuning process, images are generated by using both the fine-tuned model and the base model with the same prompt. As seen on the

Table 4.15, the outputs of the fine-tuned model exhibit clear relevance to the given prompt, whereas the base model's outputs does not demonstrate such a correlation.

Table 4.15 Comparison of the base model and fine-tuned model

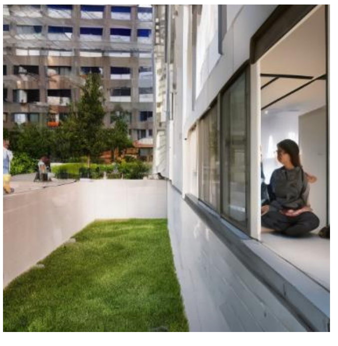
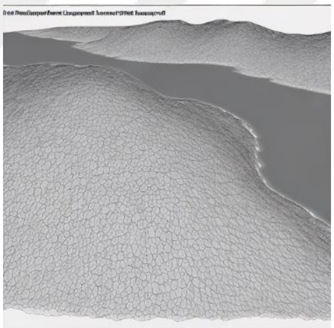
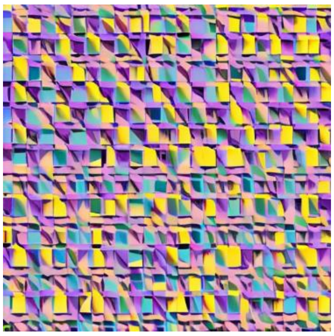
Prompt	Base Model	Fine-Tuned Model
aspect_0_process_Family_and_Multigenerational_Living		
aspect_1_meaning_The_Human_Condition_and_Existential_Reflection		
aspect_2_exterior_Privacy_and_Social_Interaction		

Table 4.15 (continued)

aspect_3_material_Exposed_Structural_Elements		
aspect_4_program_Kitchen_as_Socializing_Area		
aspect_5_spatial_Deck-Extended_Outdoor_Living		
aspect_6_region_High-Density_Urban_Suburb_		

Table 4.15 (continued)

aspect_7_location_Rapid_Urbanization_of_Villages		
aspect_8_cladding_Stained_Wood_Siding		
aspect_9_relation_Sloped_and_Embedded_Topography		
aspect_11_volume_Abstract_Cubic_Prismatic_Forms		

4.3 Correlation of Training Images and Generated Images

In order to assess the extent to which certain features in the input images influence the generated images, and to demonstrate this relationship in general, comparisons are made between the generated outputs and the original images filtered using the same parameters.

Additionally, this test helps to understand how the model responds based on the specificity of the prompt. With unclear boundaries between design research and unique design in the context of generative AI, this investigation highlights the point at which research transitions into design initiation.

Table 4.16 Comparison of randomly selected training images and generated images of forest houses in south europe and north europe regions

<table>
<tr><td colspan="2">Prompt: image_type_7_Forest_Houses_with_Simple_Forms_and_Natural_Materials, Reg_south_europe
Sample Count: 811</td></tr>
<tr><td>Random Training Images
</td><td>Generated Images
</td></tr>
</table>

Table 4.16 (continued)

Prompt: image_type_7_Forest_Houses_with_Simple_Forms_and_Natural_Materials, Reg_north_europe **Sample Count:** 616	
Random Training Images 	Generated Images 

In Table 4.16, stereotypical forest houses generated for Southern Europe and Northern Europe regions are showcased. At first glance, the color palette differences are noticeable, with lighter colors featured in the Southern Europe images, and darker colors in the Northern Europe images. However, given the generic nature of the prompts, it's challenging to assert that the generated images accurately represent the architectural characteristics of a respective region.

In Table 4.17, images are generated from prompts varying from a single to multiple parameters, moving from general to specific. As prompts become more detailed and matching projects decrease, the generated outputs align more closely with a narrowed-down example pool. However, as prompt 4 illustrates, generated images aren't strictly 'filtered' by set parameters. For example, even though the only matching project is characterized by its brute concrete material, the generated images also incorporate elements from other projects sharing some defined parameters like the common stone color of southern European projects. Notably, only when a specific project is mentioned, as in Table 4.18, are the generated images strongly limited by the features of that particular project.

This experiment highlights the contrast between the precise nature of deductive reasoning that exemplified here by filtering, and the intuitive, holistic approach inherent to neural networks. It's essential for users to recognize this fundamental difference and to understand that the generated content, rather than being an exact reflection of the input parameters, is shaped by a multitude of influences, some of which might not align with the user's original intent.

Table 4.17 Comparison of input images with generated images with varying prompt specificity

<table>
<tr><td colspan="2">Prompt 1: “image_type_2_Urban_Garden_Architecture_with_Street_Façade” (12,900 matching images)</td></tr>
<tr><td>Training Images
</td><td>Generated Images
</td></tr>
</table>

Table 4.17 (continued, 2/4)

Prompt 2: “image_type_2_Urban_Garden_Architecture_with_Street_Façade, reg_south_europe (1700 matching images)	
Training Images	Generated Images
	

Table 4.17 (continued, 3/4)

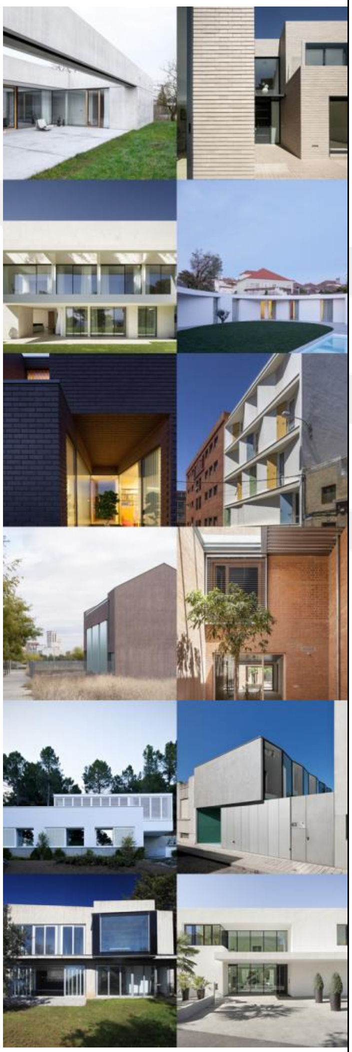
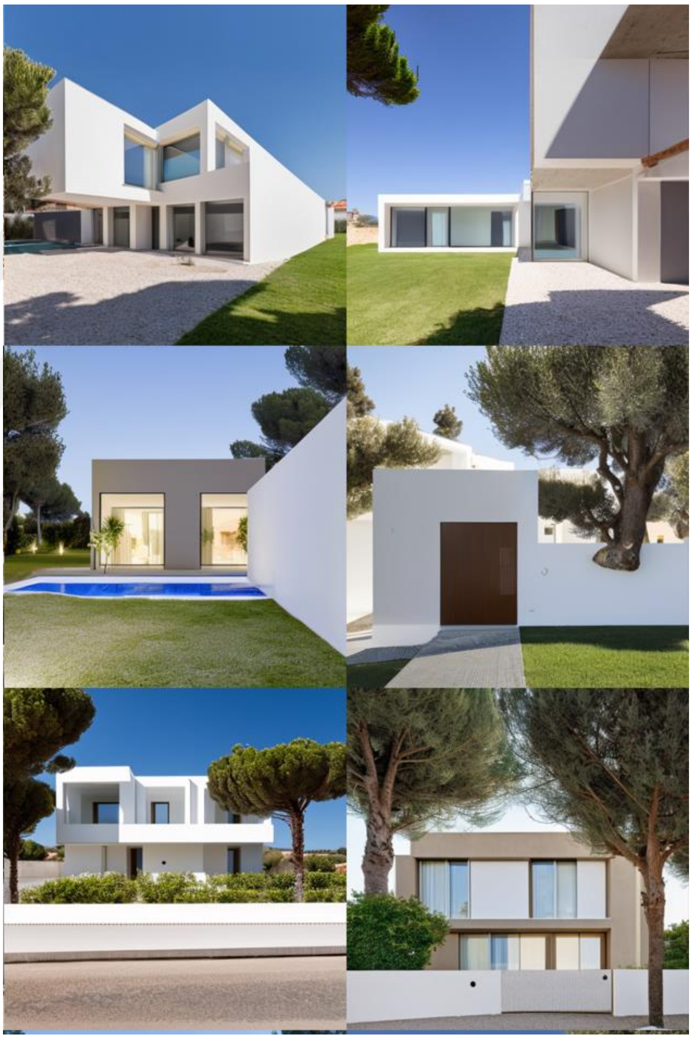
<table>
<tr><td colspan="2">Prompt 3: “image_type_2_Urban_Garden_Architecture_with_Street_Façade,
reg_south_europe,
aspect_1_meaning_Spatial_Perception “
(25 matching images)</td></tr>
<tr><td>Training Images
</td><td>Generated Images
</td></tr>
</table>

Table 4.17 (continued, 4/4)

<table>
<tr><td colspan="2">Prompt 4: “image_type_2_Urban_Garden_Architecture_with_Street_Façade,
reg_south_europe,
aspect_1_meaning_Spatial_Perception ,
aspect_3_material_Exposed_Structural_Elements”
(3 matching images, 1 project)</td></tr>
<tr><td>Training Images

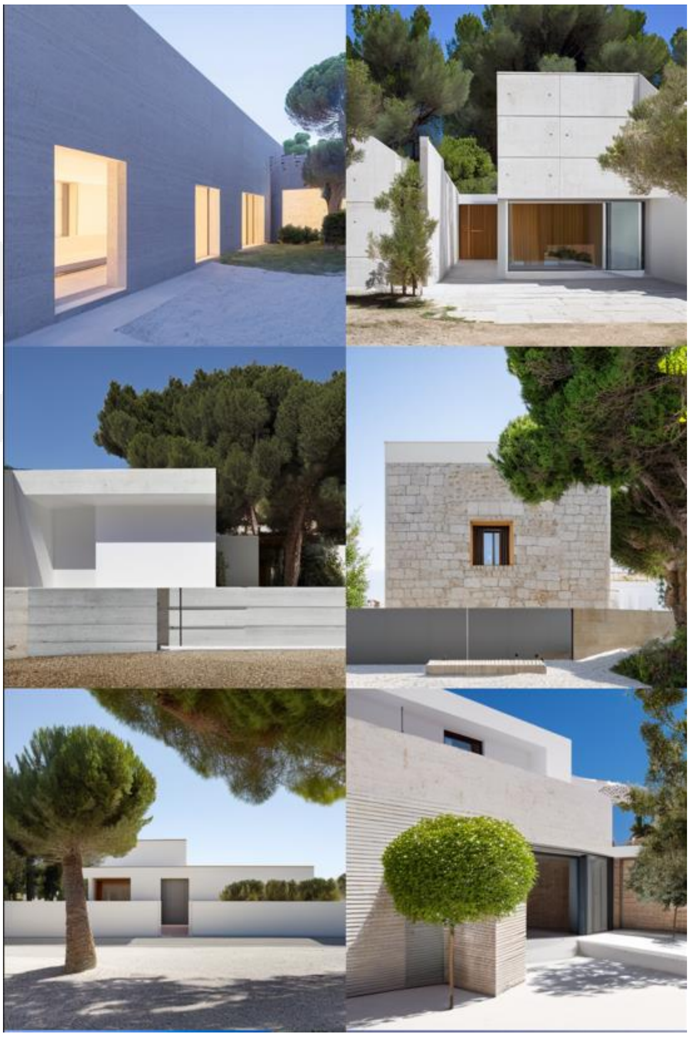
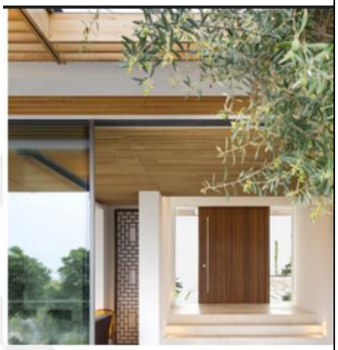

</td><td>Generated Images
</td></tr>
</table>

Table 4.18 Comparison of the input image and generated image when input project is stated in the prompt

Prompt	**Base Prompt:** city_size_3, rain_1, temp_3, reg_south_europe, image_type_2_Urban_Garden_Architecture_with_Street_Façade, aspect_1_meaning_Human-Centric_Landscape_Integration, aspect_2_exterior_Geometric_Facades, aspect_14_construction_Prefabricated_and_Modular, **Variant:** project_identifier_<...>		
Input Data (1 project)			
Generated - project specified in the prompt			
Generated - project is not specified in prompt			

4.4 Architectural Relevance and Alignment

The evaluations in this section predominantly aim to address the initial research question concerning the extent to which architectural characteristics can be extracted from architectural data, and how well-aligned the model is in this regard.

Additionally, this section delves into the potential roles and workflows of generative AI, which forms the second research question. These potential roles are proposed in this study as a continuous spectrum extending from design research to unique design creation. In this scenario, a user could employ this model for design research by exploring existing projects and understanding the common design responses of previous architects based on specific design factors.

The tests executed by generating images by using a consistent 'seed' and set of parameters, with the variation of altering a single parameter for each iteration. Each modification to a parameter provided insights into how that particular architectural characteristic is represented by the model, thereby offering an understanding of its alignment with the intended architectural design aspects.

4.4.1 External Parameters

The external parameters consist of the visual type (refer to section 3.4), project size, and a collecetd socio-geographical data based on the project location. The data collected includes the region (sub-continent), average temperature, average rainfall, city population, and the country's wealth gauged by GDP per capita. The continuous parameters were transformed into categorical ones utilizing a percentile-based discretization algorithm (refer to section 3.1),

For instance, in Table 4.19 an architect might utilize this model to discern climate related building characteristics they should consider while designing a building situated within a small town (city_size_1), with client requests emphasizing privacy around the garden area (aspect_2_exterior_Connected_and_Private_Gardens). In this context, the average temprature represent an external parameter, while privacy emerges as a parameter extracted from the architects' descriptions.

The table 4.21 shows that, as the average temprature ('tmp_0' to 'tmp_1' phrases in the prompt) increases, the openings of the generated buildings tend to increase, while the vegetation around the building also changes accordingly. This suggests that, the model captured a design trend related to the average temprature of projects' location.

Table 4.19 Variations by average temprature

<table>
<tr><td colspan="4">Prompt 1: size_3, city_size_1, image_type_6_Landscape-Integrated_Houses_with_Open_Interiors_and_Private_Courtyard, aspect_2_exterior_Connected_and_Private_Gardens, aspect_3_material_Sustainable_Timber_Cladding</td></tr>
<tr><td colspan="4">Prompt 2: size_4, city_size_2, wealth_4, image_type_2_Urban_Garden_Architecture_with_Street_Façade, aspect_2_exterior_Native_Vegetation-Rich_Landscapes, aspect_6_region_Urban_Boundary-Defined_Plots
Variant: temp_(...)</td></tr>
<tr><td>Avg Tm</td><td>Prompt 1, Seed 1</td><td>Prompt 1, Seed 2</td><td>Prompt 2, Seed 3</td></tr>
<tr><td>tmp_0 (-10 °C to 10 °C)</td><td></td><td></td><td></td></tr>
<tr><td>Tmp_1 (10 °C to 12 °C)</td><td></td><td></td><td></td></tr>
</table>

Table 4.21 (continued)

Tmp_2 (12 °C to 15 °C)			
Tmp_3 (15 °C to 19 °C)			
Tmp_4 (19 °C to 30 °C)			

As Table 4.22 clearly reveals, project size is accurately captured by the model, and it is a parameter that can be modified. While the variations in average temperature in Table 4.21 exemplify the design research use-case, the modification of project size exemplifies the use-case of generating a unique design. Users can modify or iterate towards a design artifact they aim to generate, the model can be utilized for both understanding design trends and facilitating unique design creation.

Table 4.20 Variations by project size

Base Prompt: rain_2, temp_2, reg_north_america, image_type_6_Landscape-Integrated_Houses_with_Open_Interiors_and_Private_Courtyard,
Variant: size_<...>

size	Seed 1	Seed 2	Seed3
Size_0			
Size_1			
Size_2			
Size_3			

Table 4.22 (continued)

<table>
<tr><td>size_4</td><td></td><td></td><td></td></tr>
</table>

Tables 4.23 exemplify another design research use case, where an architect utilizes the model to discern cultural or regional trends in architecture, rather than making a design decision. The results suggest that a generative model can be employed to summarize a large amount of visual data, aiding in the understanding of overall trends and cultural influences.

Table 4.21 Courtyard of a large house / Variations by Region

<table>
<tr><td colspan="3">Base Prompt: size_4, image_type_6_Landscape-Integrated_Houses_with_Open_Interiors_and_Private_Courtyard, aspect_2_exterior_Volumetric_Composition
Variant: region_(...)</td></tr>
<tr><td>South Europe
</td><td>North Europe
</td><td>West Europe
</td></tr>
</table>

Table 4.23 (continued)

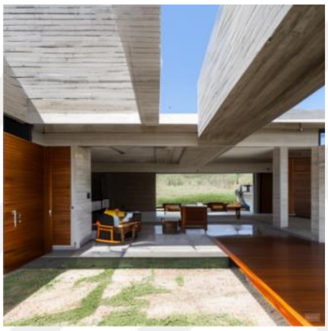
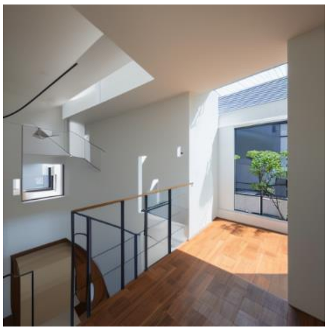
East Europe	West Asia	East Asia
		
North America	South America	South East Asia
		

Table 4.22 Minimalist, historical Interior with Garden View / Variations by Region

Base Prompt: size_4, temp_0, image_type_0_Modern_Minimalist_Interiors_with_Garden_View, aspect_0_process_Building_history_and_renovation, **Variant:** region_<...>		
South Europe	North Europe	West Europe
		

Table 4.24 (continued)

East Europe	West Asia	East Asia
		
North America	South America	South East Asia
		

4.4.2 Parameters from Topic Modeling

The parameters extracted from topic modeling, as presented in section 4.1, are mainly grouped into three clusters based on their semantic similarities. The first group, 'Design Process', (Tables 4.25, 4.26, 4.27) revolves around the intangible aspects of architectural design, focusing on conceptual stages, challenges, and the guiding philosophy or main ideas behind the design. The second group, 'Building Features', (Table 4.28) delves into the tangible, physical aspects of architectural design, detailing the materials used, structural features, and elements like the type of cladding or interior space arrangements. The third group, 'Cultural and Environmental Influences', (Table 4.29) has a broader perspective, examining the interaction between architectural design and its cultural, regional, and environmental contexts.

4.4.2.1 Design Process

Tables 4.25 and 4.26 showcase the model's alignment with the design factors related with design process and its intangible aspects. Assessing the relevance of these aspects in the generated images is not straightforward.

For instance, in Table 4.25, while 'human-centric landscape integration' displays an opening to the landscape, 'problem solving' presents a simple and clean image, and 'multi-generational living' features a chimney and a terrace-like space.

Table 4.23 Variations by Design Process and Requirements

Human-Centric Landscape Integration 	Harmony of Humans and Surroundings 	Spatial Sequence 
Multi Generational Living 	Problem Solving 	Architectural Regulations 

On the other hand, Table 4.26 suggests some correlations; examples of 'spatial sequence' consistently exhibit a sense of movement, direction, and transition between spaces, 'harmony of humans and their surroundings' exhibit a sense of stillness and openness, and 'the human condition and existential reflection' exhibits a sense of stillness and closeness. The clarity of such abstract concepts' relevance is not as clear as with the external parameters. However, there could be an intuitive correlation, which might not be easily discernible through simple observations.

Table 4.24 Variations by Meaning and Experience

Spatial sequence	Harmony of humans and their surroundings	Human Condition and existential reflection
		
		
		

Table 4.27 also illustrates the relevance of the 'design process' group, yet it diverges from Tables 4.25 and 4.26, which fall under the 'people' axis of this main cluster. The 'conceptual design' group in Table 4.27 exhibits more tangible and observable architectural design ideas. For instance, in the first prompt, the 'visual barriers' parameter led to the generation of semi-transparent eaves, 'sunlight and shadow play' resulted in a glass eave, 'privacy and social interaction' transformed the space into a more closed, private living area, and 'geometric facades' changed the openness into a wall with vertical claddings.

Table 4.25 Variations by design process / conceptual design

Base Prompt 1: size_4, city_size_5, image_type_6_Landscape-Integrated_Houses_with_Open_Interiors_and_Private_Courtyard, **Variable:** aspect_2_conceptual_design		
Sunlight and Shadow 	Privacy and soc. Intract. 	Native Vegetations 
Volumetric Composition 	Visual Barriers for privacy 	Geometric Facades 

Table 4.27 (continued)

Base Prompt 2: size_4, reg_south_america, image_type_6_Landscape-Integrated_Houses_with_Open_Interiors_and_Private_Courtyard, **Variable:** aspect_2_conceptual_design		
Sunlight and Shadow 	Privacy and soc. Intract. 	Native Vegetations 
Volumetric Composition 	Visual Barriers for privacy 	Geometric Facades 

4.4.2.2 Building Features

While the relevance of the parameters in the 'Design Process' group necessitates intuitive and less clear observations, Tables 4.28 and 4.29 exhibit a directly observable correlation between tangible building features and the generated images. This capability of the model suggests that, with enhanced control mechanisms, it could be employed for unique design creation by modifying and iterating towards a design artifact that a user aims to generate.

Table 4.26 Variations by materials and textures

<table>
<tr><td colspan="3">Base Prompt: size_2, city_size_4, wealth_2, rain_1, temp_3, reg_south_america, image_type_2_Urban_Garden_Architecture_with_Street_Façade, aspect_4_program_Flexible_Program
Variable: aspect_3_material_...</td></tr>
<tr><td>Stone Cladding
</td><td>Timber Cladding
</td><td>Monochromatic Color
</td></tr>
<tr><td>Brick
</td><td>Exposed Str. Elements
</td><td>Wood Decking
</td></tr>
<tr><td colspan="3">Base Prompt: image_type_6_Landscape-Integrated_Houses_with_Open_Interiors_and_Private_Courtyard,
Variable: aspect_3_material_...</td></tr>
<tr><td>Recycled Materials
</td><td>Timber Cladding
</td><td>Exposed Brick Masonry
</td></tr>
</table>

Table 4.27 Variations by volumetric composition

Base Prompt: image_type_7_Forest_Houses_with_Simple_Forms_and_Natural_Materials, **Variant:** aspect_2_conceptual_design		
Geometric Facade 	Sloped Roof 	Abstract Prismatic Forms 
Flexible Arrangements 	Deck Extended Outdoor 	Veranda 

4.4.2.3 Cultural and Environmental Influences

In table 4.28, different from the external parameter of region, regional and cultural context is provided by the architects. These examples reflect the explicitly mentioned regional influences.

Table 4.28 Variations by the regionan and cultural context

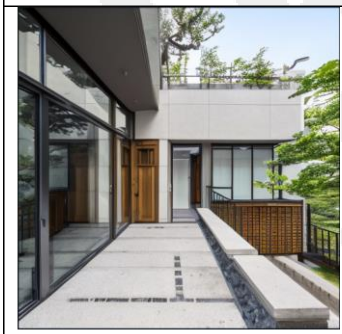
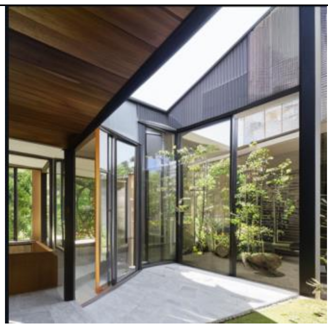
Base Prompt: image_type_6_Landscape-Integrated_Houses_with_Open_Interiors_and_Private_Courtyard, size_4, **Variable:** a aspect_6_region_<...>		
Korean Hanouk Townhouses	Japanese Residential Renovation	Vietnamese Urbanization
		
		
		

Table 4.29 Variations by the relationship of building with its surroundings

<table>
<tr><td colspan="3">Base Prompt: size_3, reg_south_europe,
image_type_4_Coastal_Houses_with_Panoramic_Views
Variable: aspect_7_location_<...></td></tr>
<tr><td>Beachside Relaxation
</td><td>Panor. Mountain View
</td><td>Lakefront Orientation
</td></tr>
<tr><td>Urban Contemplation
</td><td>Wilderness Camp. Cabin
</td><td>Urban Boundary
</td></tr>
</table>

CHAPTER 5

CONCLUSION

Artificial Intelligence has influenced all aspects of our lives over the past decade. Recently, generative AI has emerged rapidly, showcasing previously unimaginable capabilities. The foundational pursuits of the AI field have been revolutionized by machines that can proficiently handle natural language and imitate human intuition. As computational boundaries have extended beyond mechanistic domains into the intuitive and creative realms of life, the role of digital tools in design disciplines is undergoing transformation. Previously, digital tools were restricted to executing clearly defined instructions. Now, they can propose new design alternatives, support design decisions, and shift the designer's focus.

This thesis delved into these opportunities and limitations in the context of architectural design, which is a semantically rich field that can benefit from a shift towards non-deterministic, intuitive computational approaches. Architectural design is an inherently complex process that cannot be explicitly programmed with conventional methods. So far, computers in design were restricted to specific computations, knowledge management or representation of design artifacts. Today, generative AI became a promising tool that architects can employ to explore design spaces and swiftly iterate and refine their design concepts. AI's capacity to respond the design intentions of the users can expedite the design process and substantially lessen the cognitive load on the designer, especially during ideation stage. Yet, the opaque and unpredictable nature of generative AI, while offering these possibilities, also inhibits users from fully harnessing its potential. Users cannot determine if the generative AI aligns with their design intentions or understands architectural design nuances.

5.1 Summary of the Findings

In this research, the aforementioned problem caused by model opacity has been addressed by introducing an architecture-specific semantic structure to a generative model that typically lacks any kind of structure and consists only of neural network weights. In doing so, the text-to-image generative model functioned like a multi-class, multi-label conditional generative model. The topic modeling phase (refer to section 3.3) and structured captions (refer to section 3.5.1) made it possible to identify architectural design parameters that have similar impacts in the feature space. This ensured systematical observations of the models behavior.

One key takeaway is that the ability of state-of-the-art large language models to quantify the semantic similarities of documents can aid in understanding patterns in unstructured data and distilling structured and coherent knowledge. This was accomplished by applying neural topic modeling to a vast corpus of architectural text, consisting of descriptions of projects written by their designers. The distilled structure of concepts has been subjectively investigated, and the observed coherence has been described in Section 4.1. It's found that factors like data curation, selected hyperparameters, and the design of the neural network architecture all play pivotal roles. Therefore, achieving absolute neutrality in AI-based knowledge generation is not possible, as the models inevitably show a bias toward certain styles or perspectives. Instead, recognizing intrinsic subjectivity of such AI generated perspectives is imperative for safe use of this technology.

Building on the aforementioned architectural concept extraction phase, another state-of-the-art technique, latent diffusion models, has been explored. These models, in this context referred to as "stable diffusion", can generate images based on the provided text input. This was made possible by mapping the semantically meaningful vector representations of sentences (sentence embeddings) to the visual feature vector space of the image generator. In essence, different modalities have become quantifiable and therefore computable. While image generation had been achieved with Generative Adversarial Networks a decade earlier, there wasn't a way to effectively control the image generation. The existing conditional generation methods were based on image labels. However, rigorously labeling a vast number of images in a manner allowing users to generate as they desire was not feasible. With latent diffusion models, this limitation has been overcome. Yet, the success of language models and diffusion models in overcoming labeling limitations has led to entirely unstructured and uncurated models trained on vast amounts of visual and textual data unknown to the end-users. This is the central issue of this thesis and has influenced the methodology based on introducing structure to a diffusion model.

The fine-tuning process of the stable diffusion model with the Low Rank Adaptation method demonstrates that foundational models like stable diffusion and Bert can be distilled with curated data and domain expertise (refer to Section 3.5.2). This offers two benefits. Firstly, as outlined in section 4.3, it provides users the capability to examine the sources influencing a model's generation. Although it's not feasible to gauge the overall success rate of a generative model or outline a logical mechanism, this study suggests that explainability can be achieved by creating transparent models that allow users to investigate and evaluate themselves. It's also concluded that by recognizing the inherent unpredictability of generative AI and remaining alert to its risks and limitations, these systems can become invaluable assets.

5.2 Future Directions

The aim of this research was to discover and demonstrate the behavior of a text-to-image generative model in in response to an array of architectural design criteria, as well as the new form of interaction with computers. For a more comprehensive understanding and actionable insights for the development of better generative AI tools, a broader and more diverse group of human participants should be integrated into the evaluation process. This would counteract the limitations of the present study, which is grounded in insights from a singular perspective.

REFERENCES

Betzalel, E., Penso, C., Navon, A., & Fetaya, E. (2022). *A Study on the Evaluation of Generative Models*. http://arxiv.org/abs/2206.10935

Bielski, J., Langenhan, C., Ziegler, C., Eisenstadt, V., Petzold, F., Dengel, A., & Althoff, K.-D. (2022). *THE WHAT, WHY, WHAT-IF AND HOW-TO FOR DESIGNING ARCHITECTURE Explainability for auto-completion of computer-aided architectural design of floor plan layouting during the early design stages*.

Bolojan, D., Vermisso, E., & Yousif, S. (2022). *IS LANGUAGE ALL WE NEED? A QUERY INTO ARCHITECTURAL SEMANTICS USING A MULTIMODAL GENERATIVE WORKFLOW*.

Borji, A. (2018). *Pros and Cons of GAN Evaluation Measures*. http://arxiv.org/abs/1802.03446

CARPO, M. (2017). *The Second Digital Turn: Design Beyond Intelligence*. The MIT Press. https://doi.org/10.2307/j.ctt1w0db6f

Chaillou, S. (2020). ArchiGAN: Artificial Intelligence x Architecture. In *Architectural Intelligence* (pp. 117–127). Springer Nature Singapore. https://doi.org/10.1007/978-981-15-6568-7_8

Chatzikonstantinou, I., & Sariyildiz, I. S. (2017). Addressing design preferences via auto-associative connectionist models: Application in sustainable architectural Façade design. *Automation in Construction*, *83*, 108–120. https://doi.org/10.1016/j.autcon.2017.08.007

Chen, R. T. Q., Li, X., Grosse, R., & Duvenaud, D. (2018). *Isolating Sources of Disentanglement in Variational Autoencoders*. http://arxiv.org/abs/1802.04942

Croskerry, P. (2009). A Universal Model of Diagnostic Reasoning. In *Academic Medicine* (Vol. 84, Issue 8, pp. 1022–1028). Lippincott Williams and Wilkins. https://doi.org/10.1097/ACM.0b013e3181ace703

Dino, I. G. (2012a). Creative design exploration by parametric generative systems in architecture. *Metu Journal of the Faculty of Architecture*, *29*(1), 207–224. https://doi.org/10.4305/METU.JFA.2012.1.12

Evans, J. S. B. T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology*, *75*, 451–468.

Gero, J. S., & Tyugu, E. (1994). *Formal Design Methods for CAD, Proceedings of the IFIP TC5/WG5.2 Workshop on Formal Design Methods for CAD, Tallinn, Estonia, 16-19 June, 1994}*.

Goel, A. (2022). Looking Back, Looking Ahead: Symbolic versus Connectionist AI. *AI Magazine*, *42*(4), 83–85. https://doi.org/10.1609/aaai.12026

Goodfellow, I., Bengio, Y., & Courville, A. (2017). *Deep Learning*. MIT Press.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative Adversarial Networks*. http://arxiv.org/abs/1406.2661

Goodman, B., & Flaxman, S. (2016). *European Union regulations on algorithmic decision-making and a "right to explanation."* https://doi.org/10.1609/aimag.v38i3.2741

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*. http://arxiv.org/abs/2203.05794

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, *51*(5). https://doi.org/10.1145/3236009

Harputlugil, T., Gültekin, A. T., Prins, M., & Topçu, Y. I. (2014). Architectural design quality assessment based on analytic hierarchy process: A case study (1). *Metu Journal of the Faculty of Architecture*, *31*(2), 139–161. https://doi.org/10.4305/METU.JFA.2014.2.8

Hu, E., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (n.d.). *LORA: LOW-RANK ADAPTATION OF LARGE LANGUAGE MODELS*. https://github.com/microsoft/LoRA.

Huang, J., Johanes, M., Kim, F. C., Doumpioti, C., & Holz, G. C. (2021). On GANs, NLP and Architecture: Combining Human and Machine Intelligences for the Generation and Evaluation of Meaningful Designs. *Technology Architecture and Design*, *5*(2), 207–224. https://doi.org/10.1080/24751448.2021.1967060

Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, *2017-Janua*, 5967–5976. https://doi.org/10.1109/CVPR.2017.632

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Analyzing and improving the image quality of stylegan. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 8107–8116. https://doi.org/10.1109/CVPR42600.2020.00813

Koutamanis, A. (2005). A Biased History of CAAD, The bibliographic version. *Digital Design: The Quest for New Paradigms, 23rd ECAADe Conference Proceedings*, 629–637. https://doi.org/https://doi.org/10.52842/conf.ecaade.2005.629

Lawson, B. (2005). *How Designers Think: The Design Process Demystified*.

Li, Y., Fang, C., Yang, J., Wang, Z., Lu, X., & Yang, M. H. (2017). Universal style transfer via feature transforms. *Advances in Neural Information Processing Systems*, *2017-Decem*, 386–396.

Liao, Q. V. (n.d.). *Human-Centered Explainable AI (XAI): From Algorithms to User Experiences; Human-Centered Explainable AI (XAI): From Algorithms to User Experiences*.

Maslej, N., Fattorini, L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Ngo, H., Niebles, J. C., Parli, V., Yoav Shoham, Russell Wald, Jack Clark, & Raymond Perrault. (2023). *The AI Index 2023 Annual Report*.

McCarthy, J. (1996). *Defending AI research : a collection of essays and reviews*. CSLI Publications.

McInnes, L., Healy, J., Saul, N., & Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*, *3*(29), 861. https://doi.org/10.21105/joss.00861

Minsky, M. (n.d.). *Steps Toward Artificial Intelligence*.

Mirza, M., & Osindero, S. (2014). *Conditional Generative Adversarial Nets*. 1–7. http://arxiv.org/abs/1411.1784

Molnar, C. (2022). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable* (2nd ed.). https://christophm.github.io/interpretable-ml-book/. https://christophm.github.io/interpretable-ml-book/

Nilsson, N. J. (2010). *The quest for artificial intelligence : a history of ideas and achievements*. Cambridge University Press.

Otani, M., Togashi, R., Sawai, Y., Ishigami, R., Nakashima, Y., Rahtu, E., Heikkilä, J., & Satoh, S. (2023). *Toward Verifiable and Reproducible Human Evaluation for Text-to-Image Generation*. http://arxiv.org/abs/2304.01816

Reimers, N., & Gurevych, I. (2019). *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks*. http://arxiv.org/abs/1908.10084

Samuel, A. L., & Gabel, F. (2019). *Artificial Intelligence for Games: Seminar Some studies in machine learning using the game of checkers*.

Sun, J., Liao, Q. V., Muller, M., Agarwal, M., Houde, S., Talamadupula, K., & Weisz, J. D. (2022). Investigating Explainability of Generative AI for Code through Scenario-based Design. *International Conference on Intelligent User Interfaces, Proceedings IUI*, 212–228. https://doi.org/10.1145/3490099.3511119

Wang, D., Yang, Q., Abdul, A., & Lim, B. Y. (2019, May 2). Designing theory-driven user-centric explainable AI. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290605.3300831

Winston, P. H. (1990). *Artificial Intelligence at MIT: Expanding Frontiers: Vol. MIT Press* (Third Edition).

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *Proceedings of the IEEE International Conference on Computer Vision, 2017-Octob*, 2242–2251. https://doi.org/10.1109/ICCV.2017.244