Int. J. Human–Computer Studies 182 (2024) 103163

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Automatic generation of virtual architecture using user activities in metaverse

Sun-Young Jang *, Sung-Ah Kim

Department of Architecture, Sungkyunkwan University, 2066 Seobu-ro, Suwon 16419, Korea

A R T I C L E I N F O

Keywords:
Architectural design
User behavior
Image generation
Artificial intelligence
Virtual reality
Metaverse

A B S T R A C T

Recently, architectural offices and IT companies have undertaken projects to build and sell cities in metaverses. However, it is considered inefficient to undertake such construction individually, building-by-building, using 3D modeling techniques. Fortunately, the use of artificial intelligence (AI) offers a breakthrough solution. This paper proposes a method of automatic generation of architectural spaces that is based on user activities in a metaverse and that can analyze users' perceptions of such spaces. To this end, in the present study, by referencing the theory of environmental psychology, we constructed a theoretical framework that can quantitatively analyze activities, based on which, we developed a metaverse space. Architectural space images were produced based on grammatical rules, which converted the analysis results into text, as well as a deep-learning-based text-to-image generator. Twenty participants responded to two questionnaires on their perceptions and opinions after participating in the metaverse experience and after receiving the architectural space image. Consequently, we analyzed the users' opinions on the architectural space created using the data they provided. Further, we reviewed the validity of the architectural design of the virtual space created using the proposed approach and explored the possibilities of its future utilization.

1. Introduction

The capability of artificial intelligence (AI) is advancing rapidly. Recently, an AI-generated painting won an award in an art contest (Roose, 2022), and AI has expanded into various fields of writing including newspaper articles, advertising copy and fiction (Ouyang et al., 2022). The field of architecture is no exception in this regard. Several studies utilizing AI in architectural design already have been conducted, having adopted various approaches such as automatic conversion of two-dimensional (2D) to three-dimensional (3D) shapes (Park and Kim, 2021; Auliaramadani et al., 2020), indoor object searching and furniture arrangement (Chaillou, 2020; Song et al., 2022) as well as floor planning (Sharma et al., 2017; Wu et al., 2019; Nauata et al., 2021). The aim was to develop deep learning application techniques for specific tasks or steps in a series of architectural design processes. In these studies, AI was used in a limited manner as a tool to assist human architects in real-world tasks. However, in a virtual space, AI is expected to play a more active role and with greater autonomy than in a real space. Architecture in a virtual space is structurally free, as it is not physically constrained by gravity. In addition, more creative design is possible given the absence of limitations such as safety and huge construction costs.

With the recent increase in the demand for metaverses, virtual architectural spaces are gaining interest apace. The concept of a metaverse has been defined in many studies and is continuing to evolve (Davis et al., 2009; Dionisio et al., 2013; Wang et al., 2022). To summarize, a metaverse, as supported by various technologies such as big data, AI, and blockchain, seamlessly connects the real and virtual worlds. Accordingly, it can be considered to be a new piazza where value creation and communication are possible through social, economic, and cultural activities. Virtual spaces already have been implemented and used, even before the advent of metaverses. However, due to the unusual conflux of circumstances represented by COVID-19, advances in immersive content technology (augmented reality (AR)/virtual reality (VR), rendering, and networking), and the development of new technologies (e.g., AI and blockchain.), metaverses have emerged rapidly and been recognized as a new industry for the future. Many organizations worldwide, such as large IT, gaming, software, and content companies, have attempted to attract consumers by developing metaverses or participating in such platforms for promotion or marketing. A similar

* Corresponding author.
E-mail address: abyme1204@skku.edu (S.-Y. Jang).

https://doi.org/10.1016/j.ijhcs.2023.103163
Received 7 June 2023; Received in revised form 22 August 2023; Accepted 28 September 2023
Available online 10 October 2023
1071-5819/

phenomenon is occurring in architecture. Zaha Hadid Architects designed Liberland, a virtual city in a metaverse, and the Bjarke Ingels Group (BIG) designed Viceverse (Niland, 2022a), a virtual office in Decentraland, based on their architectural philosophy. Architects consider the design of a metaverse as an opportunity, as it falls within their remit and core competencies in architecture (Schumacher, 2022).

Because human beings have emotional and cognitive frameworks based on physical space, they accept, as similar, a metaverse space that is created in imitation of a physical space (Kim, 2022). Therefore, to work continuously in the real world, an architectural space is required in a metaverse. Recently, architectural offices and IT companies have attempted projects to build and sell cities within a metaverse (Yao, 2022; Niland, 2022b). Creating a city and numerous buildings in a metaverse by modeling in 3D space individually, building-by-building, clearly has limitations. In this case, using AI with an algorithm for creation of an architectural space can be leveraged. AI can design architectural spaces using algorithms, and the necessary architectural data can be directly collected and processed by various methods. In particular, instead of referencing contextless architectural data from the Web, data on various activities of users in a metaverse space can be collected and applied in space design in order to create user preferences and design uniqueness. Although users only perform their own activities, such as working, playing, and communicating with others, in a metaverse space, AI can create new virtual space buildings based on analysis of their activity data. An architecture thus utilizing the activity data of a user can reflect the taste and life pattern of that user in a customized manner and provide an unusual and interesting space in the form of a unique pattern that has not previously been revealed. However, whether such a designed building or space will actually suit the preferences of the user who is the data provider is questionable.

Therefore, the present research used AI to create a new architectural space customized for users using data generated by their unconscious activities in a metaverse, and then analyzed the perceptions and opinions of the users about that space. By analyzing users' opinions on the virtual architectural space created through the data provided by them, this study examined whether the architectural design of the virtual space is valid and feasible for the future. The series of experiments performed in this study suggests the potentiality of AI architects' automatic creation of virtual architectural spaces through user learning without being bound by existing concepts and physical constraints.

2. Research objectives

This research explored automatic generation of an architectural space according to unconscious user activities in a metaverse and analyzed the perceptions and opinions of the users about the created space. The main goals and approaches were as follows:

(1) Establish a theoretical framework for analysis of the space-usage behavior of users

In order to analyze user activities and preferences in space, a theoretical framework that can define and quantify user behavior is required. In the field of environmental psychology, many studies have already been conducted on the interaction between the built environment and people. The present study established a theoretical framework based on human-environment interaction theories such as environmental probability theory, environmental psychology, and spatial studies. The theoretical framework was the basis for designing a virtual space for experiments, analyzing user space-usage behavior after experiencing the virtual space, and formulating grammatical rules for creating architectural images.

(2) Create virtual architectural space images using user activities

User behavior data in the metaverse space was collected. The use of previously developed platforms for experiments could have been considered. However, this study required the design of the intended space and a controlled experimental environment, because it needed to analyze data on the user's movement and space preference in the virtual environment. Therefore, the virtual space experienced by users (i.e., the test participants) was implemented using the Unity game engine. Based on the theoretical framework established in the previous stage, we designed the virtual space by reflecting, in detail, the analytical elements of the participants' space utilization. The virtual space had the form of a residential space. In the experimental stage, the participants experienced this game-type virtual space and then responded to the first questionnaire. This questionnaire focused on the participants' perceptions of the metaverse and their opinions of the virtual space they had experienced. The analysis of the space preference and utilization method of each participant was based on the virtual space experience and the responses to the first questionnaire. The results of this analysis were converted to sentences according to the grammatical rules for virtual space creation. The extracted sentences were then converted into an architectural space in the form of an image. The deep learning model for this architectural image generation utilized Dall•E and Midjourney, which are currently gaining interest as text–image generation models.

(3) Analyze user (data provider) perceptions and opinions of the automatically generated architectural space

The participants responded to a second questionnaire on their thoughts and feelings about the architectural space images generated by the deep learning model. These questions used information rates devised by Mehrabian and Russell to measure and characterize complex stimuli in the environment. Finally, this study analyzes users' preferences and opinions on architectural space images by synthesizing the contents of the first and second questionnaires, and considers the automatic generation of architectural space by AI in the metaverse.

Regarding subject and scope, this study was limited to the virtual reality world among the metaverse types. The term metaverse was first used in Neal Stephenson (2008) to refer to a world where virtual and real worlds interact and create value. As research on metaverses has rapidly increased in recent years, many attempts have been made to more fully explain and define it; however, as the concept continues to evolve while reflecting social phenomena and technological developments, there is still no complete consensus. Indeed, as the scope of metaverses is wide and continuously expanding, various definitions and similar concepts exist (Park and Kim, 2022). For example, a metaverse can be classified according to the method by which the real and virtual worlds are combined. The Acceleration Studies Foundation, a metaverse-technology research group, classifies metaverses into four types: augmented reality (AR), lifelogging, mirror, and virtual worlds (Smart et al., 2007). Virtual worlds are virtual spaces that show virtual content and are accessed through computers, mobile device screens, and VR devices. Representative VR metaverse platforms include Minecraft, ZEPETO, Fortnite, Gather, and Roblox. And since many ready-made platforms have already been commercialized, a method of utilizing them for research experiments could also be considered. However, since existing platforms would have limitations in reflecting the measurement factors set in this study and in collecting data, we separately produced and utilized a controlled experimental environment.

3. Literature review

3.1. Theories on relationship between built environment and human behavior

A user's space-usage behavior is a response to a given environment and reflects their feelings and thoughts about the space (Proshansky, 1978). In this research, we aimed to determine the space preferences of users by analyzing their unconscious space usage, and the preference results, therefrom, were used to create architectural images. In this section, a review of the theories on the mutual influence of an architectural space and humans as well as the relevant previous studies on the environment–behavioral relationship is provided. The results of the review were used to establish, as will be outlined in Section 4, a theoretical framework for analysis of user space-usage behavior.

Theories on the mutual influence of an architectural space and humans can be classified into architectural determinism, environmental possibilism, and environmental probabilism, according to the perspective (Fisher et al., 1984). Architectural determinism considers that a built environment governs the behavior of users. However, environmental possibilism, which is the opposite of architectural determinism, considers the role of humans more actively and views humans as subjects who select and use an environment. Environmental probabilism is a compromise between architectural determinism and environmental possibilism, and considers that there is a causal probability between architectural design and human behavior. From a viewpoint similar to environmental probabilism, many studies have shown that human behavior, social phenomena, and psychological phenomena are influenced by a built environment and its design (Evans and McCoy, 1998; Næss, 2016; Goldhagen and Gallo, 2017; Canter, 1974).

As for other theories that explain how the environment is causally connected to human behavior, there are Barker's (1968) behavior settings theory and Gibson's (1979) affordance theory. Barker's behavior settings are one of the widely cited concepts concerning the environment. This theory holds that various units of the environment can be characterized by specific behaviors required of occupants, and can be explained by a direct link to people and their behavioral repertoire (Perkins et al., 1988). Gibson's theory of affordance refers to the possibilities of action that exist in the relationship between the actor and the environment.

Kaplan and Kaplan's (1989) preference framework and Mehrabian and Russell's (1974a) approach-avoidance theory investigated the preferences and emotional responses of people to an environment based on environmental psychology. They presented insights into human interactions with the environment. The preference framework is based on an analysis of the way people perceive and react to their environment, and is centered on information. On the other hand, the approach-avoidance theory relates to the influence of the environment on emotional states and behaviors.

In this study, among the theories on the influence between the environment and human behavior, we used the above-noted preference framework and approach-avoidance theory to analyze the space-usage behavior of the participants. In addition, because this study focused on user reactions and preferences in a virtual environment, it drew implications through a case study utilizing these environmental psychology theories in a virtual environment.

3.1.1. Preference framework

Stephen and Rachel Kaplan conducted a series of studies on environment and human preferences in the field of environmental psychology. They argued that a broadly interpreted environment significantly influences human cognition, behavior, and happiness. In particular, humans have a strong demand for information, and their cognition and emotions have a close relationship (Kaplan and Kaplan, 2009). Their findings provide insights into how people perceive and react to their environments, particularly with a focus on information. In their proposed preference framework, information is the key to organism survival, and so considering a landscape in terms of information will help to explain these preferences (Kaplan, 1975). In addition, information that facilitates understanding of an environment will be particularly prominent, they argued, and such knowledge will be closely related to environmental preferences (Kaplan, 1973).

By conducting preference response experiments for photographic materials, Rachel Kaplan revealed that indicators, which are capable of entering an environment, obtaining information, and maintaining orientation, are consistently important (Kaplan, 1985). Thus, rapid and largely unconscious determination of preferences involves evaluation of the glimpsed space and its qualities, which is strongly influenced by the potential to function in that setting (Kaplan, 1985). The results of their study showed that landscapes were preferred in cases of a high possibility of wayfinding, the presence of elements that invite a user to go deeper into the scene, and legible landscapes.

In summary, according to Kaplan and Kaplan's studies, humans show preferences for an environment in terms of functions such as obtaining information. Human preference reactions can be considered to be the results of unconscious evaluation of human needs, goals, and environmental suitability. Visual resource management tasks, Rachel Kaplan emphasize, should focus on interpreting spatial characteristics in terms of human function in relation to visual evaluation (Kaplan, 1985). Such an interpretation and perspective can be particularly useful when analyzing the results of questions asked about the feelings and perceptions of a visual environment and images.

3.1.2. Approach–avoidance theory

Based on environmental psychology, Mehrabian and Russell focused on emotional responses to psychological and social environments and identified the interaction between humans and an environment by measuring people's emotions about that environment. They proposed a theory that physical or social stimuli in an environment directly affect the emotional state of a person, thereby influencing his/her behavior (Mehrabian and Russell, 1974a). Specifically, they posited pleasure, arousal, and dominance (PAD) as emotional responses and states, and conducted experiments to characterize approach–avoidance responses to an environment. The PAD emotional response variables were taken to represent the characteristics of an environment that trigger emotions. These variables were used as parameters to determine human approach–avoidance behaviors such as physical approach, task performance, exploration, and social interaction (Mehrabian and Russell, 1974a).

The PAD theory has been referenced in many studies as a background theory and/or applied for measurement of reactions and emotions. Based on it, the behaviors and emotions of people in various situations and spaces have been analyzed. Some examples are behavior of people in shopping spaces (Robert and John, 1982), experiences of visitors in art galleries of hospitals (Fang et al., 2012), and acceptance of a technology by consumers (Kulviwat et al., 2007). However, because the present research aimed to analyze the preferences of people for an environment, Mehrabian and Russell's experimental process and research framework were the reference foci.

To support their hypothesis, they tested participants' approach–avoidance responses to environments. In these experiments, participants were provided 40 situations in the form of texts and asked to respond on a factor analysis scale (Table 1) of approach–avoidance and state–anxiety responses. As shown, this factor analysis scale consists of

Table 1
Results of mehrabian and Russell's factor analysis of approach–avoidance and state–anxiety responses to 40 situations (Mehrabian and Russell, 1974a).

Factor	Item
Factor 1: Preference for the situation	1. Liking of
	2. Preference for
	3. Tendency to seek
	4. Desire to explore
	5. Desire to leave
Factor 2: Desire to affiliate with the situation	6. Desire to be friendly
	7. Desire to initiate conversation
Factor 3: Desire to solve problems in the situation	8. Desire to solve difficult problems
	9. Desire to solve easy problems
Factor 4: State anxiety	10. Tense feeling
	11. Nervous feeling
	12. Anxious feeling

nine approach–avoidance items (three factors) and three anxiety items (one factor). Factor 1 characterizes the change in preference for a situation. It consists of liking, preference, tendency to seek, and desire to explore or leave the situation. Factors 2 and 3 characterize affiliative and problem-solving responses, respectively. Factor 4 consists of adjective pairs about the states of anxiety, ease, and nervousness. The scale used in the experiments showed that specific behavioral responses to situations were related. The scale was effectively used to support Mehrabian and Russell's initial assumptions that characterized general preference responses. Therefore, in order to analyze the preference response, we constructed a theoretical framework based on the scale used in the course of this experiment.

3.1.3. Application of environmental psychology theory in virtual environment

Since the advent of virtual environments, environmental psychology theories have been used to study human activities, reactions, and relationships in such environments. In this section, a review of studies that have analyzed user responses by using the approach–avoidance theory and preference frameworks in the virtual environment is presented. The theory of Mehrabian and Russell can be used to analyze preference responses to situations and environments, and in fact, it has been referenced in many studies. In particular, it has been used to investigate human behavior and emotion in the real and virtual worlds, and had been confirmed to be valid.

Clark et al. (2009) investigated the impact of website design characteristics on the emotions, attitudes, and behaviors of users. Moreover, they used the approach–avoidance theory to apply, to a physical environment, website-environmental insights into the relationship between emotions and behaviors. They considered the approach–avoidance theory suitable for study of the design characteristics of websites. They argued that research on consumer behavior in physical spaces supports the importance of PAD emotions influencing consumption, and that the same results are supported for website applications.

Citing the hypotheses of Mehrabian and Russell, Huang (2003) showed how the reactions and emotions of users toward the information load in a virtual environment affect their search and shopping in such a space. However, in order to apply the approach of environmental psychology to virtual online shopping, modification of the theoretical framework and review of shopper behavior are necessary so as to properly account for the characteristics of the virtual environment. Consequently, Huang demonstrated that both information and emotion play important roles in determining whether users approach or avoid a given environment. Örer (2016) investigated various influencing factors in the interior space design of a virtual environment and compared the emotional reactions of subjects to that environment. The positive and negative emotional responses of the subjects experiencing virtual spaces designed with three different shapes—straight lines, symmetrical curves, and asymmetrical curves—were compared.

Yüncü (2015) conducted a virtual-environment study based on the preference framework of Kaplan and Kaplan. The relationship between the virtual destination environment on the Internet (My Destination Barcelona Facebook page) and satisfaction and loyalty was investigated. According to the research results, if a virtual destination environment includes various visuals and sufficient information as well as environmental factors that stimulate curiosity, excitement, and entertainment, it can provide satisfaction and loyalty to visitors. Demangeot and Broderick (2010) developed and tested a Gestalt model for how consumers perceive an online shopping environment. The authors conceptualized the main perceptual attributes of the online shopping environment and modeled their relationships by analyzing the preference framework of Kaplan and Kaplan and studies related to online shopping. Their study showed that people perceive an online shopping environment as an environment that they want to understand and explore, despite its being a virtual environment. They also were able to confirm that the preference framework of Kaplan and Kaplan could, in fact, be used in an online shopping environment.

In summary, previous studies using the approach–avoidance theory and the preference framework have demonstrated that they are theoretically valid not only in the physical world but also in a virtual environment. These studies measured user actions, reactions, and emotions in virtual spaces and investigated their preference responses to information and searches. Therefore, we refer to these theories as the basis for analyzing preference responses to a virtual environment and explaining the derived results.

3.2. Image generation using deep learning

3.2.1. Text-to-image generation

A text-to-image model (i.e., a text-to-image generator) is a generative AI model that understands a natural language and outputs images when a natural language is used as the input. Since the introduction of generative adversarial networks in 2014 (Goodfellow et al., 2014), which generate new images based on learned images, the development of models that generate images from text in computer vision has significantly progressed. In January 2021, Dall•E (Ramesh et al., 2021) was released by OpenAI, after in 2022, Dall•E 2[1] was released, high-performance large models such as Midjourney, Stable Diffusion, and Imagen (Saharia et al., 2022) appeared consecutively. These the latest models all utilize a diffusion model. Sohl-Dickstein et al. (2015) suggested an initial concept for creation of a flexible and tractable diffusion model by systematically and slowly destroying the data structure by an iterative forward diffusion process and subsequently learning a reverse diffusion process that restores the data structure. Since then, in the area of generative modeling, this diffusion process has provided impressive generation functions for various purposes such as image generation and high resolution, inpainting, and editing (Croitoru et al., 2023).

In this research, we performed architectural image generation using Dall•E and Midjourney, powerful image generation models that provide different styles for the generated image. Dall•E has two stages of operation: a prior one that creates a contrastive language-image pre-training (CLIP) image embedding given a text caption, and a decoder that creates an image based on the image embedding (Ramesh et al., 2022). Dall•E's text and image embedding utilizes CLIP (Radford et al., 2021) developed by OpenAI. CLIP is a contrastive learning model that matches images and text prompts. CLIP minimizes loss of photorealism and caption similarity and improves image diversity by learning a powerful image expression that captures both semantics and style (Ramesh et al., 2022). Stable Diffusion is also based on CLIP. As for Midjourney, the technical mechanism of the model is undisclosed.

With large-scale models emerging in succession, studies testing their performance and usability also have been increasing in number. Borji (2022) argued that if the images generated by Dall•E and Midjourney are good, a deep model should be able to recognize them. Borji input such a synthesized image into an object recognition model to measure the classification accuracy. Witteveen and Andrews (2022) analyzed text prompts to utilize diffusion models such as Dall•E, Midjourney, and Stable Diffusion. A text prompt is an input written in a natural language that instructs a text-to-image model to generate an image. Witteveen and Andrews determined words and phrases likely to belong to a specific category, and analyzed the impact of these words and phrases on the images produced. Abduljawad and Alsalmani (2022) presented an approach for creating a remote-sensing dataset, which is difficult to collect using an image generation model. They suggested methods to generate specific datasets by comparing and utilizing image generation models in scenarios without extensive datasets.

[1] Dall·E mentioned later means Dall·E 2.

3.2.2. Studies using text–image generation in architecture

Various fields such as art, design, business, and medical care are showing interest in using image generation models. The urban and architectural fields are also presenting similar trends. Studies using models such as Dall●E, Midjourney, and Stable Diffusion are emerging.

Seneviratne et al. (2022) argued that there is a need for a method to accurately capture the semantic information present in complex scenes, such as built environment scenes. They evaluated the potential of large-scale pre-trained text of image generation models to capture a built environment in terms of health, safety, and aesthetics. They evaluated the ability of Dall●E to address issues of urban design and planning, the biases present in its image-generating functions, and its ability to capture complex concepts such as healthy and safe design. They created queries by which to classify the four dimensions of urban design (image format, geography of generated images, urban quality, and urban units). In conclusion, the study found that Dall●E has considerable potential to assist human experts in this field with domain- and task-specific datasets.

Ploennigs and Berger (2022) argued that diffusion-based AI models could be powerful tools in the initial architectural design phase, which includes the ideation, sketching, and modeling sub-stages. They investigated the applicability of Dall●E, Midjourney, and Stable Diffusion to architectural design, and analyzed 40 million query datasets of Midjourney to determine how it uses natural language processing. Consequently, they developed a workflow for interior and exterior design by mapping specific design tasks to each model. The workflow uses Midjourney, which is excellent for free-form ideas, at the idea stage in order to create a desired scene, and subsequently materializes it through Dall●E and Stable Diffusion to obtain a specific scene.

Yildirim (2022) observed that AI could expand the spectrum of creative capabilities by enabling designers to explore more design ideas from new perspectives. Yildirim compared the Dall●E, Midjourney, DiffusionBee, and MotionLeap models, and examined the possibility of using text-to-image models in architecture, interior, and urban design. In particular, in architectural design, a text-to-image model will be the most applicable one in the initial design stage and will be helpful in visually expressing design concepts rapidly and in iterating and exploring various options.

As reviewed, the related studies have various viewpoints and goals for model utilization, and variables are set differently according to those utilization goals. Table 2 summarizes examples that use the text-to-image model for studies in architecture and urban areas. It highlights the following aspects: the goal of model utilization, the design stage to be applied, the model used, and the parameters for text and image generation that were analyzed. These parameters are the items that are adjusted for output of results within the target and range for model utilization. Adjustment of parameters, therefore, can directly affect image results.

Since high-performance text-to-image models have been released relatively recently, there are still relatively few studies that have used them in the field of architecture. The relevant studies have used models mainly to explore ideas and creative designs in the early stages of architectural design. It also should be noted that parameters vary according to the utilization goal of the model. Key parameters include the format of images (style variants such as posters and cartoons), the image processing functions of models (editing, upscaling, etc.), items related to architectural expression (floorplan, façade material, etc.), and items related to scale or unit (gradual scale of interiors, architecture, cities, etc.; urban units such as suburbs, streets, etc.).

Overall, the above-mentioned studies indicated that the intuitive and fast image generation function of the text-to-image model has significant potential in the work stage wherein concepts of architectural design are visualized. Moreover, studies have commonly said that models can be used efficiently and creatively in the design process. On the other hand, studies have pointed out that the text-to-image model needs to consider the appropriate use of natural language, because when using it, wrong images can be generated.

We also investigated how previous studies using the text-to-image model set conditions and created images, and considered which parameters they employed. Most studies have used models to quickly generate creative results in the initial planning stage or idea exploration stage. Regarding the setting of parameters, studies have set different classifications of parameters according to the purpose of their utilization. However, it was confirmed that for a large category, parameters

Table 2
Text-to-image model and parameters used in previous studies in field of architecture.

Authors (year)	Goal for use	Design stage	Model (Generator)	Parameters for text and image generation
Seneviratne et al. (2022)	Assist human experts in the built environment and urban planning.	Planning and design phase	Dall·E	A query was created by combining four relevant dimensions of urban design: image format, geography, urban quality, and urban unit.
Ploennigs and Berger (2022)	Provide support for ideation, sketching, modeling, etc. This is intended to match and utilize models suitable for specific design stages.	A series of early design phases (from idea to implementation)	Dall·E, Midjourney, Stable Diffusion	Experimentation with in-/out-paint, editing, upscaling, and semantics for items of ideation, sketches, collages, building variants, image combination, style variants, construction plans, exterior design, interior design, and creation of textures.
Bolojan et al. (2022)	Address the limitations of language-based models by providing interconnected AI-assisted workflows. Streamline machine-vision-related tasks for architectural design.	Design stage	Dall·E, VQGAN+CLIP, CLIP-Guided Diffusion Models	As centered on semantic explanations such as building, plan, section, and façade, sentences are transformed and experimented on with more specific forms such as "a modernist glass façade."
Yildirim (2022)	Explore design ideas quickly and diversely from a new perspective.	Ideas and early design stages	Dall·E, Midjourney, DiffusionBee, MotionLeap	Evaluate the possibility of utilization when the scope of application is set as architectural, interior, and/or urban design application.
Paananen et al. (2023)	Support creativity for development of students' architectural design concepts.	Early design phase	Midjourney, Stable Diffusion, Dall·E	The design theme (culture center) and site are given in common, and students' concepts (e.g., floorplan, interior perspective visualization, facade material sample) are expressed visually.
Ploennigs and Berger (2023)	Quantitatively analyze 58 million queries, and discuss whether the results of the most popular queries (words and expressions) related to architecture can be applied to architectural design.	Ideation, architectural collages, building variants (planning and design phase)	Midjourney	The input text was constructed using the most frequently used queries, and words corresponding to content words, style words, and quality words were included in one sentence, respectively.

can be classified into items related to image format, the image processing function of the model, architectural expression, scale, and unit.

4. Methodology

4.1. Research pipeline

As established in previous studies, using a text-to-image model for architectural space design has strengths in intuitively and quickly creating creative images in the initial design stage. However, it cannot be said that architectural design is simply a matter of easily outputting new and creative images. It is important to design the architectural space considering the behavior, characteristics and opinions of the user, the subject of use. In that respect, the user's behavior or response data vis-a-vis the space can be a key idea in creating a unique design considering a specific user.

The present research aimed to analyze the user's activity in exploring spaces in the metaverse environment and, using that analyzed data, to create an architectural space image through a text-to-image model. In addition, since there is no existing study that has investigated users' opinions and reactions to architectural space images created in this way, we analyzed users' preferences and opinions through experiments and surveys. In order to achieve the stated goal, the following complex tasks were carried out: 1) analysis of user behavior in space, 2) production of a metaverse space wherein experiments are possible, 3) generation of architectural space images using text-to-image models, 4) survey conducting to collect user opinions. The necessity and considerations for concretely setting the method of performing the proposed tasks are as follows.

- Analysis of user behavior in space: Analysis of user behavior in space should consider the mutual influence of the architectural space and people. Research on this has long been carried out in the field of environmental psychology. Since the advent of virtual space, traditional theories on the mutual influence of architectural space and people have been studied to see if they are applicable in virtual space as well. In the present research, it was necessary to refer to studies that have investigated people's feelings and preferences for the environment in order to quantify and capture the characteristics of users' reactions and behaviors to space.
- Production of a metaverse space wherein experiments are possible: A metaverse space for experiments should be produced based on a suitable configuration and plan to discriminate user behavior. How to discriminate and interpret user behavior is an important issue. Discrimination and interpretation of user behavior are based on the theory of environmental psychology, as noted above. To facilitate discrimination and interpretation, the same criteria should be applied to the design of a metaverse. Therefore, it is necessary to create a metaverse experience space that reflects discriminatory factors and is controllable, rather than a commercialized metaverse space.
- Generation of architectural space images using text-to-image models: The performance of deep learning models differs depending on the amount of training data. Large, commercialized models such as Dall•E and Midjourney train models for a long time using vast amounts of data, which is difficult for individuals to do. Since these latest models can obtain high-quality images if the logic for input text composition is refined elaborately, we employed them in the image output stage using AI.
- Survey conducting to collect user opinions: In order to investigate the user's reaction to the generated image, a process of listening to opinions through a survey is required. In particular, it is necessary to ask questions about users' specific feelings and thoughts about images, and about the applicability of AI architects in terms of methods. The results will be able to provide meaningful implications and design considerations for the creation of architectural spaces using AI.

Prior to the detailed description of each part of the methodology, this section provides an overview of the entire methodology used in this study. In addition, the system configuration as organized to obtain the research goals is presented. Fig. 1 provides the system structure indicating the flow of data, the configuration of modules, and overall operation.

The metaverse space is an experiential space in which users participate, and is designed to conduct experiments. The fabrication of the metaverse is described in detail in Section 4.2.2. The metaverse space composition and arrangement of objects are deliberately designed for extraction of user behavior data. The criteria for this design follow the reference module. The reference module consists of a theoretical framework for analysis of the user's space-use behavior and parameters representing the characteristics of the space. The reference module was established based on the theories of environmental psychology dealing with the interaction between humans and the environment, and its detailed composition is shown in Section 4.2.1.

The operation process of the system centers on user participation as follows: ① Users freely experience the metaverse in the form of a virtual space. ② Each user's spatial experience data is stored in the repository under the user experience date. ③ The user responds to the first survey immediately after the experience. In the first survey, questions are asked about the existing perception of the metaverse and the metaverse space experienced (Section 4.4). This response data is stored in the first survey data repository. ④ The user's experience data and survey response data are comprehensively analyzed in the analysis module. The analysis

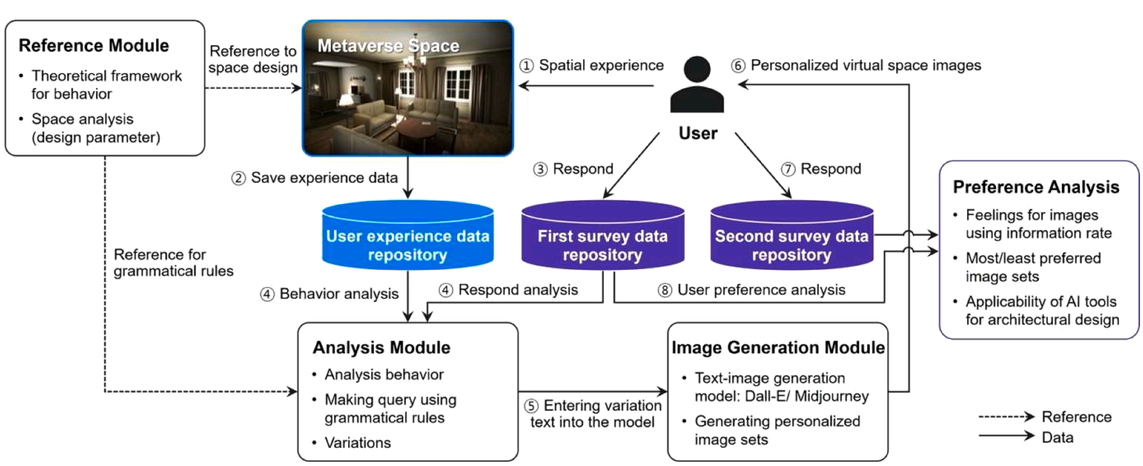


Fig. 1. System architecture.

result becomes a query in the form of a sentence according to the grammatical rules set in this study. The grammatical rules used here refer to the reference module (Section 4.3.1). When one query sentence is completed, it is diversified into 6 variations to generate image sets. ⑤ Six variations of text are transmitted to the image generation module and input into the text-image model (Dall•E, Midjourney). ⑥ The created image is a personalized result, because it is based on individual space exploration and survey response data (Section 4.3.2). These images are passed back to the participants. ⑦ Participants respond to the second survey based on the images. The second survey asks questions centered on feelings and preferences for images (Section 4.4). ⑧ Based on the data of the two surveys, the user's preferred image type, feeling for the image, and applicability of the AI tool to architectural space design are comprehensively analyzed. In this study, the participants were preliminarily divided into three groups according to the degree of architectural education.

4.2. Data collection and analysis of user activity in metaverse

4.2.1. Theoretical framework for activity-data analysis

This section presents the theoretical framework that serves as the basis for the planning of a metaverse space and for the analysis of data generated after a user experiences that space. Data analysis was performed to determine the unconscious preference for the space during the space experience as well as the individual characteristics of the user. The theoretical framework was established based on a literature review, as indicated earlier, and was designed in consideration of determination and measurement that were performed in the results analysis stage after the experiments. This research adopted the viewpoint of the environmental probability theory, which considers probability as it relates to the relationship between architectural design and human behavior. It focuses on the interaction between an environment and individual variables and analyzes the respective characteristics of people and space (Porteous, 1977). In this research, we analyzed data according to those two classifications, people and space. Moreover, the characteristics of people were further subdivided as follows: 1) how people use space, 2) the characteristics of people.

The framework for the analysis of people is summarized in Table 3. This analysis consisted of general information about people's space-use behavior and items requiring their (the users') unique propensities. Obtaining general information on the space-usage behavior of users required measuring the time and frequency of their space use. Humans do not stay for a long time or frequently visit spaces that they do not prefer unless there is a special purpose in being there. Therefore, a space that was visited for a long time or frequently was assumed to reflect the preference of the corresponding user to some extent. A previous study on user space-usage behavior set usage time, frequency, purpose, and location as the main collection items (Park and Kim, 2019). In this study, because the unconscious uses of a space were measured, the purpose of use was not collected. Also, given that a location derives the name of a space according to the time and frequency of its use, the purpose of use was considered to correspond to those measures, and thus was excluded. The space usage time is the time spent in a space. We measured the time that users stayed in all spaces and determined the space with the highest value. However, due to differences in the usage time for each person, the degree of preference was adjusted such that the results were not uniform. The setting of the interval for the degree of preference also plays a role in adjusting, such that diverse variations emerge when generating images later. The preference determination according to the staying time was divided into four stages (normal, slightly strong, strong, and very strong) based on 20 s intervals. The set criterion for time was supplemented by a preliminary test conducted before this experiment to make relevant determinations without concentrating on one section. The frequency of space usage measures the number of visits to a space and determines the space with the highest value. For frequency, four steps were set to determine the degree of preference, and the reference value was corrected by a preliminary test.

In the preliminary test, the entire process involving construction of the theoretical framework, conducting of experiments, image creation, and administering of two surveys was conducted in advance. The experiments were done with three people, and the suitability of the determination and measurement standards as well as the analysis method was supplemented by referring to the process, results, and opinions of the participants.

User characteristics are referenced in order to determine the unique propensity of a user toward a space. The composition of the overall items in Table 3 was matched by referring to the items presented in Table 1 (Mehrabian and Russell, 1974a), which had been used by Mehrabian

Table 3
Developed theoretical framework for analysis of user activity data.

Category	Purpose	Explanation	Method for determining or measuring	Determined interval setting	Approach–Avoidance factor (Mehrabian and Russell, 1974a)
General information about the space usage behavior of a user	Space-usage time	Time spent in a space by each space	Measuring the time spent in each space, the name of the space with the highest value is derived (unit: s) (* Total experiment time limit)	Four levels of strong or weak preference for a space: (0 to 20) s (normal), (20 to 40) s (slightly strong), (40 to 60) s (strong), and over 60 s (very strong)	• Desire to stay or leave • Liking of (△)[1] Preference for (△)
	Space-usage frequency	Frequency of use by each space	Measuring the frequency of use for each space, the name of the space with the highest value is derived (unit: number of times)	Four levels of strong or weak preference for a space: 0 to 2 times (normal), 2 to 4 times (slightly strong), 4 to 6 times (strong), and 6 or more times (very strong)	• Preference for • Liking of (△)
User properties	Territoriality	Do users regard a specific space as their domain?	Inferring a space in which individuals feel territorial through queries that can identify behaviors in which they regard the space as their own territory	A space selected by querying placed objects; it is a space where one wants to be alone.	• Preference for (△) • Liking of (△)
	Exploration	Does a user have a tendency to explore the space?	Determine the search range in the virtual space in the form of a given village by setting the area to houses, roads, and other houses	Search only a given house (weak), search a road (medium), visit other houses (strong) → choose a path to explore based on the intensity	• Desire to explore • Avoidance of exploration • Tendency to seek (△)
	Preference for simplicity and complexity	Are users interested in exploring complex or simple spaces, or do they try to avoid them?	Determine, by the first questionnaire, whether users prefer complex or simple spatial structures	Query the spatial structure to determine whether it is complex or simple, and assign one of four levels depending on the response	• Desire to solve difficult problems • Desire to solve easy problems

[1] (△) indicates that it applies to a wide range.

and Russell for characterizing the approach–avoidance response to environments. Moreover, we reconstructed the items to suit the purpose of the analysis.

We derived territoriality, exploration, simplicity, and complexity as items that can identify spatial preferences while reflecting the items presented in Table 1. Measuring and determining the Factor 2 and Factor 4 items in Table 1, as based on user movements in a virtual environment from an observer's perspective, were difficult. Therefore, we decided to refer to the responses of the users to the first survey conducted after experiencing the metaverse.

The "territoriality" item is about whether a particular space is regarded as the domain of a user. Territoriality can be inferred based on a specific behavior that regards the space as the territory of the user. People express territoriality in many ways; among them, this study used the method of arranging symbols or placing important objects in terms of their effects on the space (Wise, 2000). In the metaverse-experience experiments and the first survey, to identify the space where territoriality is felt, certain objects that were important to the metaverse users (test subjects) were assumed. Subsequently, where the users wanted to put those objects, and the places where they wanted to stay alone, were queried.

The "exploration" item relates to whether the user tends to travel or move around the space. To determine exploration, the accessible space during the metaverse experience was divided into three parts (the houses given for the first search, road, and other nearby houses), and the intensity of the search was set. If a user explored only around the houses given for the first search and for a limited time, the user was determined to have a relatively "weak" tendency to explore; if a user looked around the neighborhood via the road in front of the front yard, the user was determined to have a "medium" tendency; if a user went into several other houses in the neighborhood and explored the inside, the user was determined to have a relatively "strong" tendency.

The "preference for simplicity and complexity" item relates to whether a user is interested in exploring complex spaces or tends to avoid them. This item has limitations in terms of setting the degree of simplicity and complexity when designing the virtual space of the metaverse for experience. Therefore, in the first survey conducted after experiencing the metaverse, we obtained participant responses to a pairwise comparison on a nine-point scale regarding whether they wanted the space they experienced to be more complex or preferred it to be simple.

Following the analysis of people, the framework for analyzing the characteristics of a space was developed, which is summarized in Table 4. This analysis determined the characteristics of a space when a specific space was selected after analyzing the characteristics and usage behavior of people. The purpose of the spatial analysis was not to identify a generally average good space but to infer the individual preference of the users for spatial characteristics. The categories that characterize a space included lighting, windows, color, furniture, and privacy. The classification of the categories was reconstructed by referring to *Behavior and elements of architectural design* by Fisher et al. (1984). Table 4 shows the categories for the analysis of spatial characterization. Such characterization was intentionally reflected in the design of the virtual space for the conversely controlled experiment, so that each space had distinctive characteristics from the beginning.

Table 4
Categories for analysis of spatial characterization (Fisher et al., 1984).

Categories of spatial characteristics	Detailed elements (design parameters)
Lighting	• Illuminance by single light intensity • Illuminance by number of lights • Lighting color
Window	• Number of windows • Sizes of the windows • Shapes of the windows
Color (Munsell, 1905)	• Achromatic color (white, black, and gray) • Cold color (blue–green, blue, and purple–blue) • Warm color (red, yellow–red, and yellow) • Neutral color (green–yellow, green, red–purple, and purple)
Furniture	• Arrangement of the furniture - Sociocentric, open, and welcoming interactions (e.g., round table) - Socially centrifugal, closed, and suppressing social contact (e.g., waiting room) • Expression of the space function through the furniture
Privacy	• Planning around "visual intrusion" - House made of glass - Single and double rooms - How to place barriers around the working space

4.2.2. *Construction of metaverse-experience environment in form of virtual space*

Among the metaverse types in which virtual and real worlds are mixed in various forms (Smart et al., 2007), this study was limited to the VR world. A previously developed platform could have been used for user experimentation; however, this research required the design of an intended space and a controlled experimental (and virtual) environment in which to analyze the data on user movements and space preferences. Therefore, a metaverse space was designed by referring to the criteria set for the experiments and the theoretical framework, and was built using the Unity game engine. Each user participated in the experiments in an operation similar to a first-person VR game (cf. First-Person Shooter (FPS)). They could walk, run, turn a flashlight on/off, and open and close doors in the metaverse.

The experimental space built in the metaverse was centered on a main house with a garden. Six similarly shaped houses were built in the neighborhood (accessible) on a road that crossed a village. A user started the experience from inside the main house. This house was the main space designed for user behavior analysis and consisted of two floors above the ground, a basement, and outdoor spaces (a swimming pool and a garden) (Fig. 2). The interior space of the house included seven living rooms (four shared and three private), four rooms including a bedroom and other functions (a private living room, study, and dressing room), four bathrooms, and two kitchens. In addition, it consisted of a laundry room, a garage, a study, a storage unit, and an entrance. This house had several spaces for the same purpose, such as seven living rooms and four bathrooms. This was to extract the distinguishing characteristics of the space, when it was later assumed to be the preferred space, by using the same or slightly different design elements of the space, such as lighting, windows, and furniture. These elements are specified in Table 4.

All of the spaces in the building had intentional planning elements based on the items listed in Table 4. The lighting had parameters of none (0), few (1), normal (2), and many (3), and the intensity and color of a lighting fixture were adjusted according to its shape. The windows were planned to have parameters of none (0), few (1), normal (1 wall: 2), and many (2 or more in 2 walls). Among the planning elements, furniture was used to express an open or closed type in the living rooms and dining spaces. The open-type furniture arrangements mainly utilized sofas and tables of various sizes in a circular or facing form. The closed-type furniture was arranged in rows. The elements of color and privacy were primarily combined in the bedroom. Privacy was categorized into single and double rooms. The four rooms included a double room in the form of a dormitory and three single rooms with a private living room (only), a private living room with a small study, and a private living room with a bathroom, respectively. In addition, the bedrooms were designed with patterns (flower patterns) and cold (blue, purple–blue, and dark blue–green), warm (bright red and yellow–red), and neutral (green–yellow, green, and light purple) colors. The objective was to characterize different atmospheres.

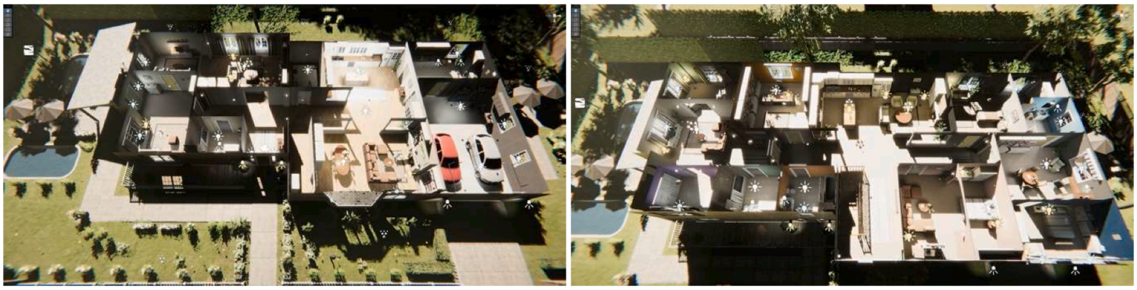


Fig. 2. Compositions of first floor (left) and second floor (right) of house constructed with Unity for metaverse experience.

4.3. Generating virtual architectural space images using user activity

4.3.1. Grammatical rules for generation of architectural images

Architectural image generation uses text–image generators based on deep learning. Therefore, converting previously collected behavioral data into text according to certain rules was necessary. After the user activity data were analyzed using the framework presented in Section 4.2.1, we established a series of steps to convert the results into sentences. In the framework for data analysis (Section 4.2.1), the characteristics of people and spaces were analyzed separately, and this classification is valid at the stage of setting rules for image creation. The order of application of the rules was as follows: 1) the spaces were derived according to the user's space-usage behavior and characteristics (Table 3); and 2) the characteristics of the space (Table 4) were extracted only for the derived space.

Table 5 lists examples of the grammatical rules and sentence conversion for generation of architectural images. First, for the "space usage time" item, the time spent in the entire space was analyzed, and the space with the highest value was derived. Even if several people stayed in the same space, the time they stayed there was different. Therefore, the intensity was set for this, and the criteria set in Table 3 were followed. The degree of preference for space was classified into four levels, which, in the query composition stage, were replaced with areas (slightly wide, wide, very wide, and extremely wide). Finally, the derived space reflected the characteristics (Table 4) given during the space design.

The second item, "space usage frequency," is the number of visits by a subject to all spaces, according to which the space with the highest frequency was derived. Space preference as inferred by frequency of use was classified into four stages according to the number of visits and substituted for area in the query composition stage. The space derived from this item also reflected the characteristics provided when designing the space, as listed in Table 4.

Finally, the three items of user characteristics (territoriality, exploration, and preference for simplicity and complexity) may have different resulting values, depending on the responses and behavior of the user.

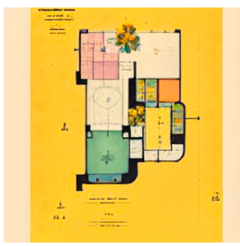
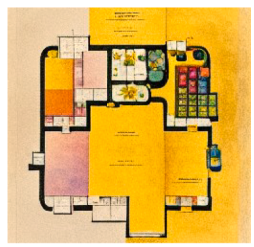
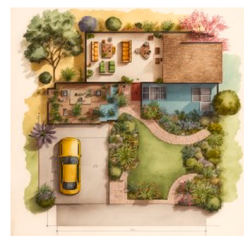
Table 5
Grammatical rules for generation of architectural images.

Category	Application content and order	Query (sentence) composition and example	
Space usage time (Duplicate possible)	① Derivation of "the longest-stayed space" - The name of the space with the highest value regardless of the length of stay	• "(name of the space)" (e.g., "living room")	very large living room with many windows
	② Degree of preference for the space - Choose from four levels: (0 to 2)0 s (normal), (20 to 40) s (slightly strong), (40 to 60) s (strong), and over 60 s (very strong)	• "(the degree of preference according to the scale)" (e.g., "very large")	
	③ Characteristics of the space - Reflecting the characteristics from the five "characteristics of space"	• "(Characteristics of space)" (e.g., "many windows")	
Space usage frequency (Duplicate possible)	① Derivation of "the space with the highest frequency of use" - The name of the space with the highest value regardless of the frequency	• "(name of the space)" (e.g., "courtyard")	medium-sized courtyard with many flowers
	② Degree of preference for the space - Choose from four levels: 0 to 2 times (normal), 2 to 4 times (slightly strong), 4 to 6 times (strong), and 6 or more times (very strong)	• "(the degree of preference according to the scale)" (e.g., "medium-sized")	
	③ Characteristics of the space - Reflecting the characteristics from the five "characteristics of space"	• "(Characteristics of the space)" (e.g., "many flowers")	
Reflect the applicable items (Duplicate possible)	• (if) "Territoriality" - The space selected through the query on "Place objects, space where you want to stay alone" → Derivation of the space name and reflection of the corresponding characteristics in "Characteristics of the space"	• "(space name and characteristics)" (e.g., a master bedroom in neutral colors with a private living room)	a master bedroom in neutral colors with a private living room, many pathways to explore, and simple structure
	• (if) "Exploration" - Search only the given house (weak), search the road (medium), visit other houses (strong) → choose how many paths to explore based on the intensity	• "(medium)" (e.g., many pathways to explore)	
	• (if) "Preference for simplicity and complexity" - Result of the query on "Simpler or more complex than the experienced space" → Select from the four levels for simple/complex according to the response (1 - moderately complex, 2 - complex, 3 - very complex, and 4 - extremely complex)	• "('simple' preference of Level 2)" (e.g., simple structured)	

Table 6
Image-expression ability test of Dall•E for floor plans.

Image Model	"Floor plan of ~"		"Colored floor plan of ~"	
Dall•E				

Table 7
Image-expression ability test of Midjourney for floor plans.

Image Model	Version 3, Quality 5		Version 4, Quality 2	
Midjourney				

Therefore, only items with the resulting values were reflected. And, because determining the degree of territoriality, simplicity, and complexity only through observation from an omniscient point of view was difficult, the responses from the first survey were reflected. Preferences for simplicity and complexity were provided on four levels—slightly simple (complex), simple (complex), very simple (complex), and extremely simple (complex)—according to the pairwise comparison nine-point scale answers of the user. For the territoriality item, the arrangement of objects to which characteristics were assigned and the space wherein the user wanted to stay were queried, and the name and characteristics of the space were derived. For the exploration item, if a subject only looked around the inside of the given house, the degree was set as "weak" and provided as the default value. In addition, when exploring by going outside the house to the road, the degree was set as "medium" to create an architectural image with "many paths to explore." In the case of entering other houses in the neighborhood and looking around the interiors, the degree was set as "strong" and converted to "very many paths to explore."

The sentence derived as an example of the query composition in Table 5 was "A simple-structured house with a very large living room with many windows, a medium-sized courtyard with many flowers, a master bedroom in neutral colors with a private living room, and many pathways to explore." The query derived from the grammatical rules in Table 5 comprised of a series of modified sentences. To compare the final architectural images, the subject (object) of the image was selected. In this research, the subject was limited to "house" and "floor plan," as types of architectural images.

4.3.2. Generation of architectural images using deep learning models

The architectural images were created using Dall•E and Midjourney. The images printed using these deep learning models were presented to the participants in the second survey. The images provided to the subjects were different, having been personalized for each user; however, the compositions of the images were sampled according to a certain standard. In this study, each model was tested several times using different tests to determine its ability to create architectural images and achieve quality for comparable composition and expression.

In the case of Dall•E, when a floor plan output was requested, the result mimicking the image style of the floor plan was shown. The grayscale images on the left of Table 6 resemble a floor plan initially; however, closer inspection reveals that there is no meaning to the functional aspect of the space, the furniture drawn inside, or the name of the space. As shown on the right side of the table, when a colored floor plan is ordered instead of a grayscale floor plan, it is easier to distinguish the space, and many factors showing that the command (input text) is better reflected can be inferred.

In the case of Midjourney, the version and quality can be specified. In particular, the version affects expressiveness given the large differences in the styles of the drawing images. Therefore, the Midjourney version was selected based on how much it reflected the content of the text. Moreover, images were output with the highest quality allowed based on the time of the experiment (November 12 to 30, 2022). Because the quality selection affected the details of an image, it was set to the highest level to adequately reflect the text. Table 7 presents the same sentence in different versions of Midjourney: "Colored floor plan of a house with a very large front yard with many flowers, a slightly larger garage with two cars parked, a yellow-walled study with closed furnishings, and many paths to navigate." Versions 3 and 4 have distinctly different styles. Both versions are non-realistic planes that do not fit the logic; however, Version 4 better reflects the original text.

The image format configuration was similar for both models. Depending on the forte of the models in terms of expressiveness, Dall•E and Midjourney were set to output high-quality photos and 3D-rendered images, respectively (Table 8). In the case of Dall•E, relatively interesting and natural images were output when specifying unique painting styles (e.g., Vapor wave and Matisse) or realistic spaces such as high-quality photos. Moreover, architecture in a virtual space is characterized by its not having to conform to the physical constraints of the real world, such as gravity. Therefore, we examined the reactions of the subjects by adding free and creative images that are unlikely to exist in the real world, and, as a painting style that can draw this type of building, an architectural space with keyword "surrealism" was added. Dall•E could express more smoothly and distinctively when designating a specific painter than when ordering a surrealist style. Therefore, the

Table 8
Image-expression ability tests of Dall•E and Midjourney for styling.

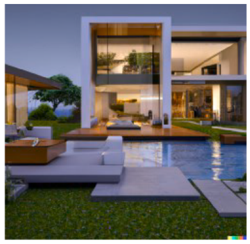
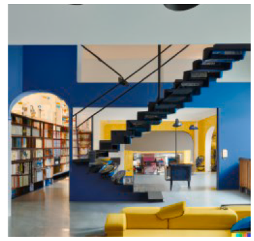
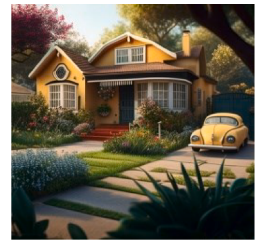
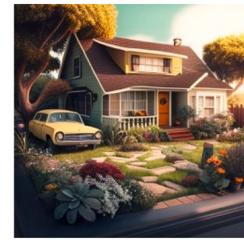
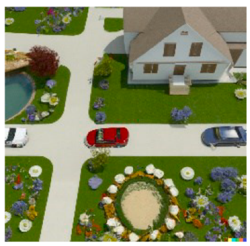
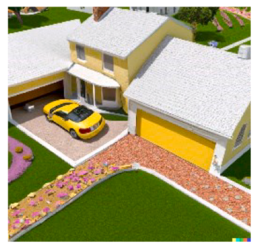
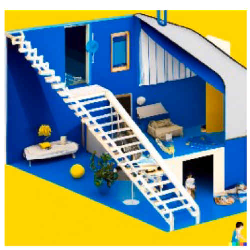
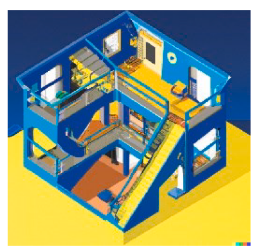
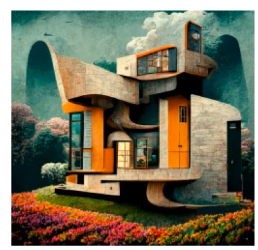
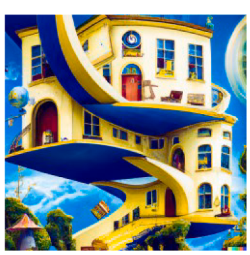
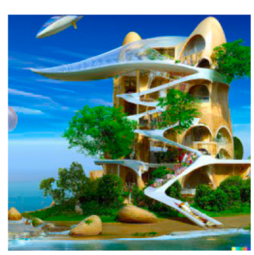
Image Style	Dall•E Image		Midjourney Image	
"High-quality photo"				
"3D rendering"				
"Surrealism style"				
"Vladimir Kush style"				

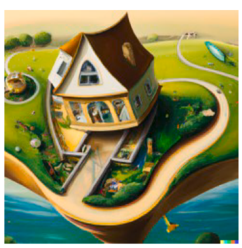
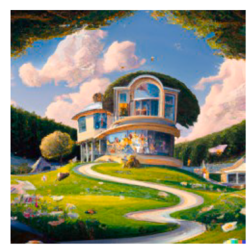
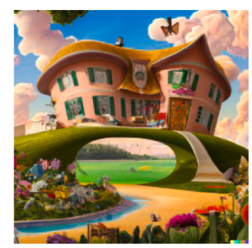
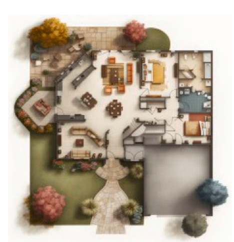
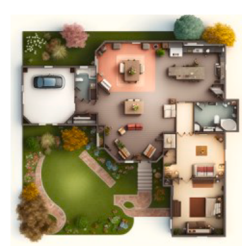
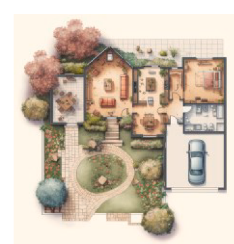
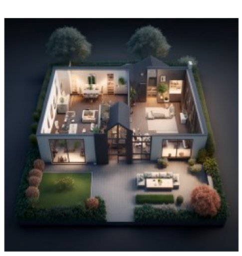
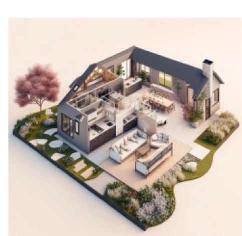
Table 9
Compositions of text variations.

Set No.	Model	Building Expression	Image Style	Version
1	Dall•E	Floor plan	Colored floor plan	-
2	Dall•E	House	High-quality photo	-
3	Dall•E	House	Vladimir Kush style	-
4	Midjourney	Floor plan	Colored floor plan	Version 4, quality 2.
5	Midjourney	House	3D rendering	Version 4, quality 2
6	Midjourney	House	Surrealism style	Version 3, quality 5

styles of various surrealist painters were substituted, and finally, the surrealist painter "Vladimir Kush style" was applied. Midjourney was effective when mandated for a "Surrealist style" than when specifying the style of a particular painter.

Table 9 presents the compositions of the image sets as the output. In the experiments, six image sets (24 individual images) were generated using the two models thrice. As the type of architectural space to be output with the image generator had been set to "house" and "floor

Table 10
Images created using Dall•E and Midjourney.

Set	Image Style	Image			
1	Colored floor plan (Dall•E)				
2	High-quality photo (Dall•E)				
3	Vladimir Kush style (Dall•E)				
4	Colored floor plan (Midjourney)				
5	3D rendering (Midjourney)				

(continued on next page)

Table 10 (*continued*)

Set	Image Style	Image			
6	Surrealism style (Midjourney)				

plan" previously, each model outputted one set of floor plans and two sets of house images.

The method of deriving an image was as follows: 1) the user behavior and characteristics were derived by analyzing the video of the activities of a user, and then the preferred spaces and their spatial characteristics were derived; 2) the text was constructed by grammatical rules; 3) to create six sets of images, the text was varied with the image format as configured in Table 9; 4) the images were created using Dall•E and Midjourney. Table 10 presents the images derived using this process.

Table 10 presents a set of images derived when the following sentence was inserted as the result for one of the 20 subjects: "(the floor plan of) a house with a very spacious master bedroom with a private living room and a purple dressing room, a rather large open plan living room with a TV and many windows, a stairway to the second floor, a storage room under the stairs, and a flowery front yard." Keeping the same sentence, only the image style was varied according to the compositions specified in Table 9. The subjects received their results in a common image set configuration, as provided in Table 10. The images and prompts used in the experiments are available at GitHub: https://github.com/DIG-SJang/virtual-architecture-image.

4.4. First and second surveys

The first survey was conducted immediately after the metaverse experience. Its purpose was to ask questions about the preferences and perceptions of each participant regarding the metaverse, their opinions on the virtual environment created for the experience, and to obtain additional information for analysis of the user in terms of territoriality, exploration, simplicity, and complexity. Table 11 presents the main contents of the first and second surveys.

The second survey was conducted to analyze the perceptions and opinions of the subjects on the presented images. The subjects (the same as those who had taken the first survey) were asked to respond to images derived by analyzing their individual activities. All 20 subjects were provided 24 different images (six sets). The second survey consisted of questions about the feelings for the image sets and individual images, as well as opinions about the automatic generation of buildings in the metaverse. In the survey, it was possible to select a set of preferred or non-preferred images directly and to provide opinions subjectively. Feelings for the image sets and the individual images were measured using Mehrabian and Russell's information rate (Mehrabian and Russell, 1974a).

Information rate was designed for a systematic, environmental psychology research approach; average information rate, for example, can be used to characterize complex spatial and temporal arrangements of stimuli in and out of an environment (Mehrabian and Russell, 1974b). According to Mehrabian and Russell, information rate can be applied to important all types of stimulus configurations, from color patches to cities, and can be used for verbally described situations, photographic or pictorial presentations, audiovisual recordings, and direct viewing by subjects. Therefore, this study used the information rate as a tool for evaluation of the feelings about images and environmental settings in the metaverse.

Table 11
Contents of surveys.

Survey	Classification	Question details
First survey	Factor 1: Basic information and opinions of the participants on the metaverse	• Basic information of the subject: age, major, and level of familiarity with virtual-environment operation • Positive and negative scales (nine points) and detailed comments about activities in the metaverse • Whether you have experience using a metaverse platform, the platform you have used, and the duration of being immersed in the virtual environment
	Factor 2: Opinions about the virtual space experienced	• The feeling for the virtual space experienced in the experiment • Questions about territoriality (Arrangement of important objects that belong to the test subject and can be used only by the subject, and the reason for it. The space where the subject would like to stay if the subject spends time alone, and the reason for it) • Other comments
Second survey	Factor 3: Opinions on the image set	• Respond, on the adjective pair scale, to how the images feel overall • Select the most preferred image set and the reason for it • Select the least preferred image set and the reason for it
	Factor 4: Opinions on the individual images	• Select the most preferred image and the reason for it and respond on the adjective pair scale • Select the least preferred image and the reason for it and respond on the adjective pair scale • Opinions on what it would be like if buildings, like images, were automatically generated in a metaverse

5. Experiments and result analysis of architectural image generation using user activities in metaverse

5.1. Experimental outline and procedure

The main purpose of the constructed metaverse was to collect the activity data of the experiment participants (metaverse users). The experimental procedure was as follows: 1) delivery of experimental guidance and precautions, 2) metaverse experience for 15 min, 3) conducting of the first questionnaire immediately after the experience, and 4) conducting of the second questionnaire by referring to the images printed individually after approximately 1 to 2 weeks.[2]

[2] Informed consent was obtained from the participants prior to the experiment.

Fig. 3. Metaverse experience scenes of user.

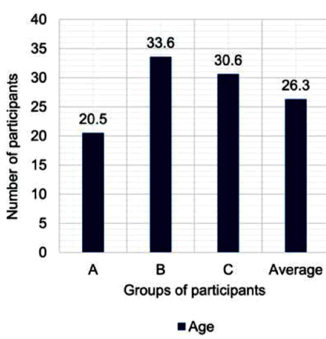


Fig. 4. Age of the participants.
* Group legend: A (Undergraduate group), B (Non-architectural majors group), C (Postgraduate student group)

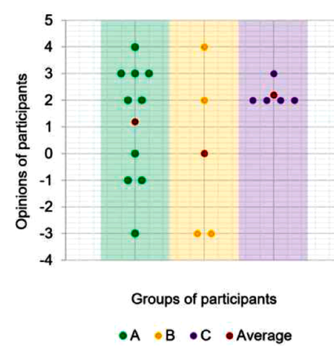


Fig. 6. Positive (+) and negative (-) opinions about the metaverse.
* Group legend: A (Undergraduate group), B (Non-architectural majors group), C (Postgraduate student group)

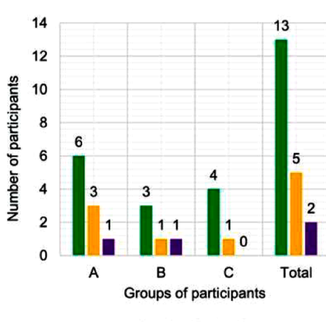


Fig. 5. Familiarity with virtual environment operation.
* Group legend: A (Undergraduate group), B (Non-architectural majors group), C (Postgraduate student group)

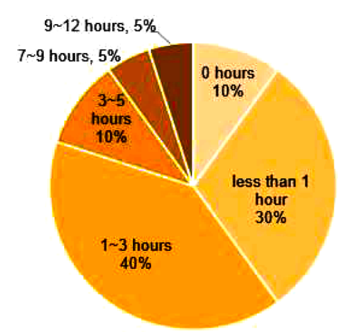


Fig. 7. How long you want to stay in the metaverse (per day).
* Group legend: A (Undergraduate group), B (Non-architectural majors group), C (Postgraduate student group)

Twenty subjects participated, and they were divided into three experimental groups: undergraduate (ten subjects), non-architectural major (five subjects), and postgraduate (five subjects). Given that in the experiments, the images created using the activities of each subject were shown again to them to elicit their response to the questionnaire, participants who could be tracked and re-answered were targeted. In terms of the experimental method, and similarly to a case study (Elor and Song, 2020) that received personalized results after subjects participated in an experiment, Elor and Song (2020) conducted an investigation using ten participants to evaluate an image in VR, generated images after learning the results using AI, and made four participants re-evaluate the personalized images. We devised experiments according to the purpose of this research by referring to the experimental processes and selection of the number of subjects in previous

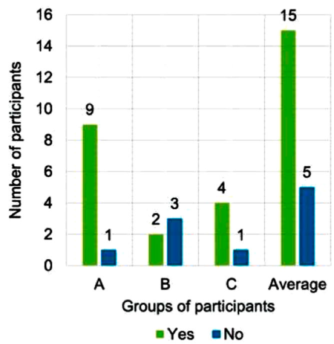


Fig. 8. Metaverse experience.
* Group legend: A (Undergraduate group), B (Non-architectural majors group), C (Postgraduate student group)

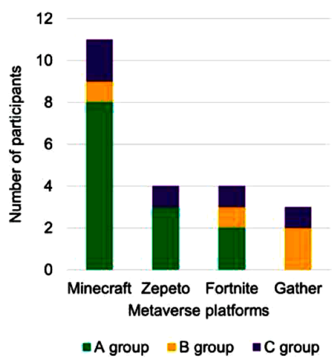


Fig. 9. Platforms you have used (experienced only).
* Group legend: A (Undergraduate group), B (Non-architectural majors group), C (Postgraduate student group)

studies. Moreover, we reflected the supplementary points found in the preliminary tests (guidance, graphic card speed, and content of response sheet).

The time limit of the experiments was 15 min. A house in a virtual environment for experimentation needs an average of 9 to 10 min during which a person familiar with the space composition and skilled in first-person games manipulates the areas in the entire space. Therefore, in this study, 15 min was set as a sufficient time to extract various behavioral results, such as exploring, staying in, or revisiting a space encountered for the first time. The experimental environment consisted of a monitor (24 inches, for transmission of the virtual environment), a keyboard, and a mouse for movement manipulation.

Prior to participating in the metaverse experience, the subjects were informed of the operation methods, the procedures for the entire experiment, the time limit, and movement recording. At this stage, the participants were asked to freely explore spaces for unconscious space exploration. However, the detailed mechanism of the experiments was left unexplained to facilitate free exploration. After the second survey was completed, the participants were informed that an image had been generated by analyzing their individual activity data, and their opinions were solicited comprehensively. The control screen viewed by the participants is shown in Fig. 3.

5.2. Experimental result analysis

- Factor 1: Basic information on participants and their opinions on metaverse

Factor 1 consisted of the basic information on the participants and their existing opinions on the metaverse. The experiments were conducted with the previously mentioned three groups. In Figs. 4–15, these groups are marked as A (undergraduate group), B (non-architectural major group), and C (postgraduate student group), respectively. The basic information for each group is as follows. Group A consisted of ten first-year students from the Department of Architecture, with an average age of 20.5 years. Group B consisted of five non-architectural major persons (majors in materials engineering, computer science, advanced materials engineering, electronic engineering, and mechanical engineering, respectively), with an average age of 33.6 years. Group C consisted of five postgraduate students majoring in architectural design, with an average age of 30.6 years. The average age of the 20 participants was 26.3 years (Fig. 4).

Fig. 5 shows the results of the degree of familiarity with virtual-environment manipulation. “Good” corresponds to being good at manipulating first-person computer games such as FPS. “Fair” describes the participant who has played first-person games before but is awkward at it. “Poor” describes a participant who has never played before. Among all of the participants, 13 had “good,” 5 had “fair,” and 2 had “poor” degrees of familiarity. Many participants responded that they were proficient. Fig. 6 plots the results of a query about positive or negative opinions on activities (socializing, watching performances, marketing, and education) in the expanding virtual environment or metaverse. This question was given on a nine-point scale to indicate “neutral (0), slightly (±1), moderate (±2), strong (±3), and very strong (±4)” for positive and negative, respectively. The overall average was “slightly positive (+1.15),” with 13 positives, 2 neutrals, and 5 negatives. The averages for Groups A–C were +1.2, 0, and +2.2, respectively, with positive responses in the order of the postgraduate, undergraduate, and non-architectural major groups.

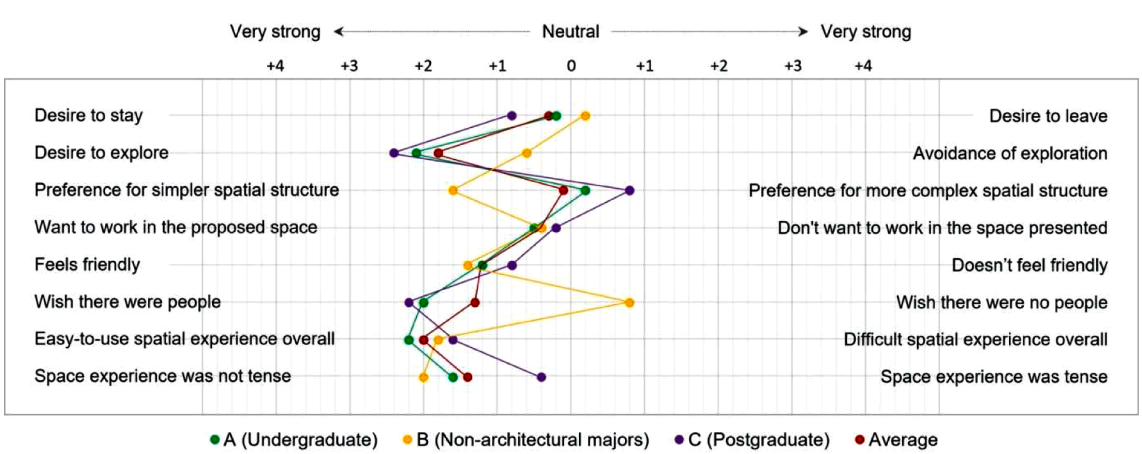


Fig. 10. Overall feeling for experienced metaverse (rearranged from question in Mehrabian and Russell’s approach–avoidance experiments).

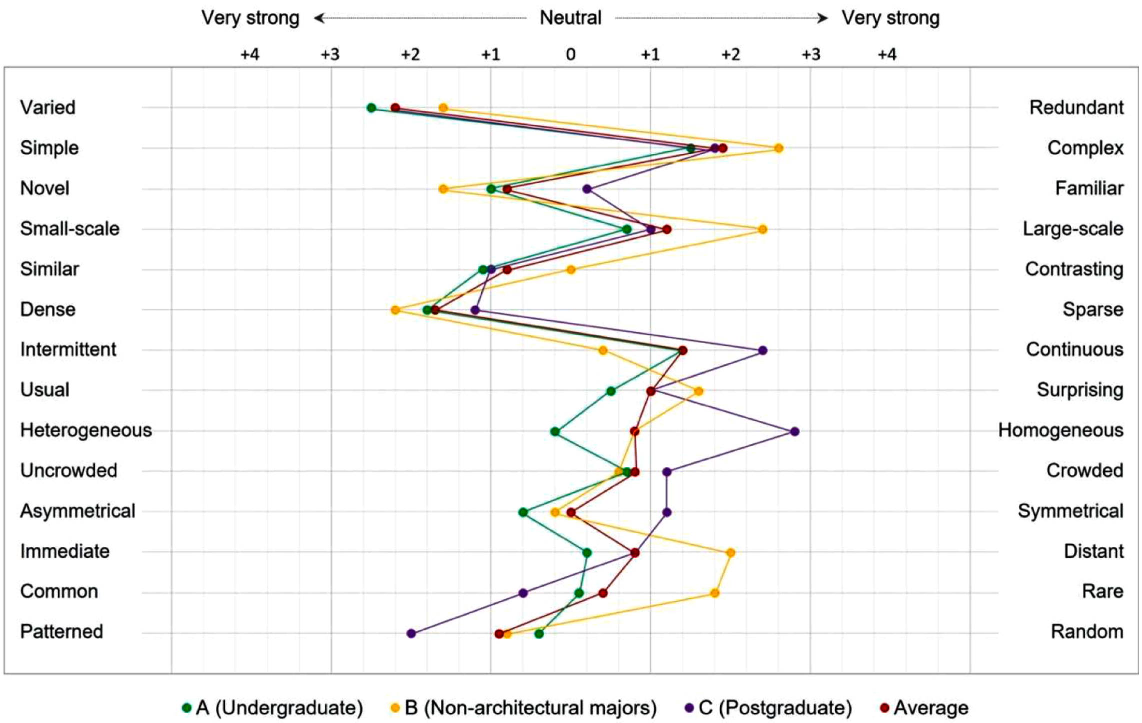


Fig. 11. Feelings for image sets (using information rate).

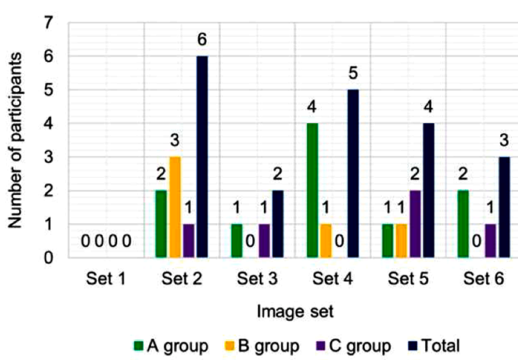


Fig. 12. Most preferred image sets.
* Group legend: A (Undergraduate group), B (Non-architectural majors group), C (Postgraduate student group)

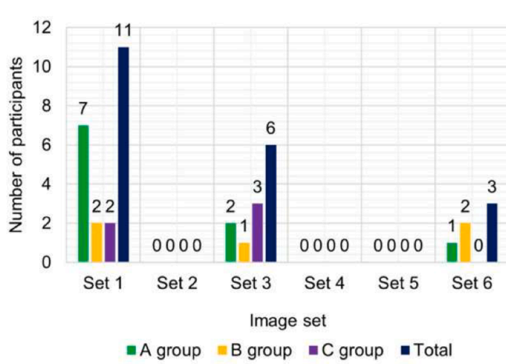


Fig. 13. Least preferred image sets.
* Group legend: A (Undergraduate group), B (Non-architectural majors group), C (Postgraduate student group)

Positive and negative opinions on the metaverse were also explained in subjective terms. Positive opinions included answers such as "efficient in terms of saving distance or time," "new attempt," and "experience something that has not been experienced." The neutral opinions were "clear advantages of both spatial and temporal freedom of VR and liveliness and health of reality" and "never thought about it in particular." The negative opinions included "there is still a big gap between virtual and reality, so it is difficult to replace" and "it is better to experience it with your own eyes." Some participants described the shortcomings they found regarding activities in schools or workplaces that had been replaced with virtual environments during the past few years of COVID-19. The main responses included that communication, such as facial expressions or accurate nuances, became inaccurate, concentration was reduced, and a sense of realism could not be conveyed.

Fig. 7 shows the responses to "how long a participant would like to stay in the metaverse during the day." Eight subjects (40 %) answered that they would stay for 1 h to 3 h, and six (30 %) answered, less than 1 h. Fig. 8 plots the responses to whether a participant had ever experienced a metaverse. Fifteen participants (75 %) responded that they had done so previously. Fig. 9 plots the responses to the questions on the platforms used by those with a metaverse experience. The most frequent response was Minecraft (11 people), followed by ZEPETO (4 people), Fortnite (4 people), Gather (3 people), and Roblox (2 people).

- Factor 2: Opinions on metaverse experience

Factor 2 consisted of questions about the subjects' feelings on the metaverse created for the experiments and the determination of territoriality. Fig. 10 plots the results of the responses to the overall feelings for the metaverse space experienced. The scale used in this question was a reconstitution of the items used in the experiments of Mehrabian and Russell, as mentioned in Section 3.1.2. The subjects were asked to mark each of the opinions in the extreme-opinion pair on a nine-point scale of "neutral (0), slightly (±1), normal (±2), strong (±3), and very strong (±4)."

Among the average response results (red graph), "I want to explore

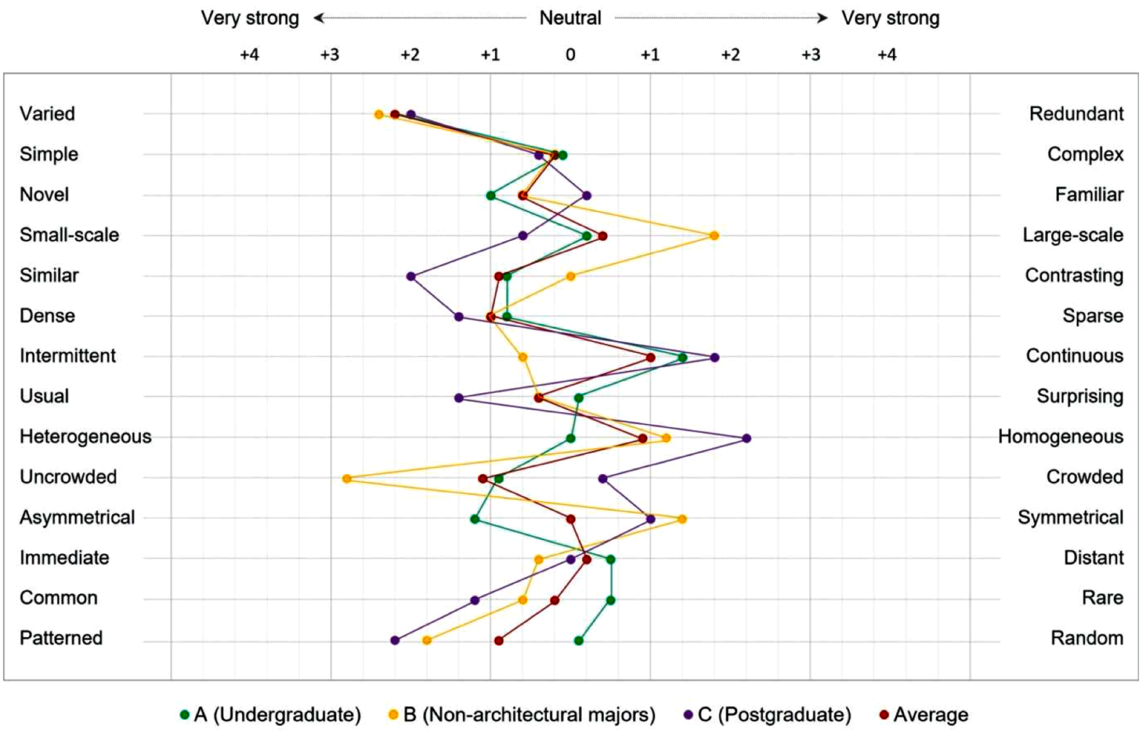


Fig. 14. Feelings for most preferred image (using information rate).

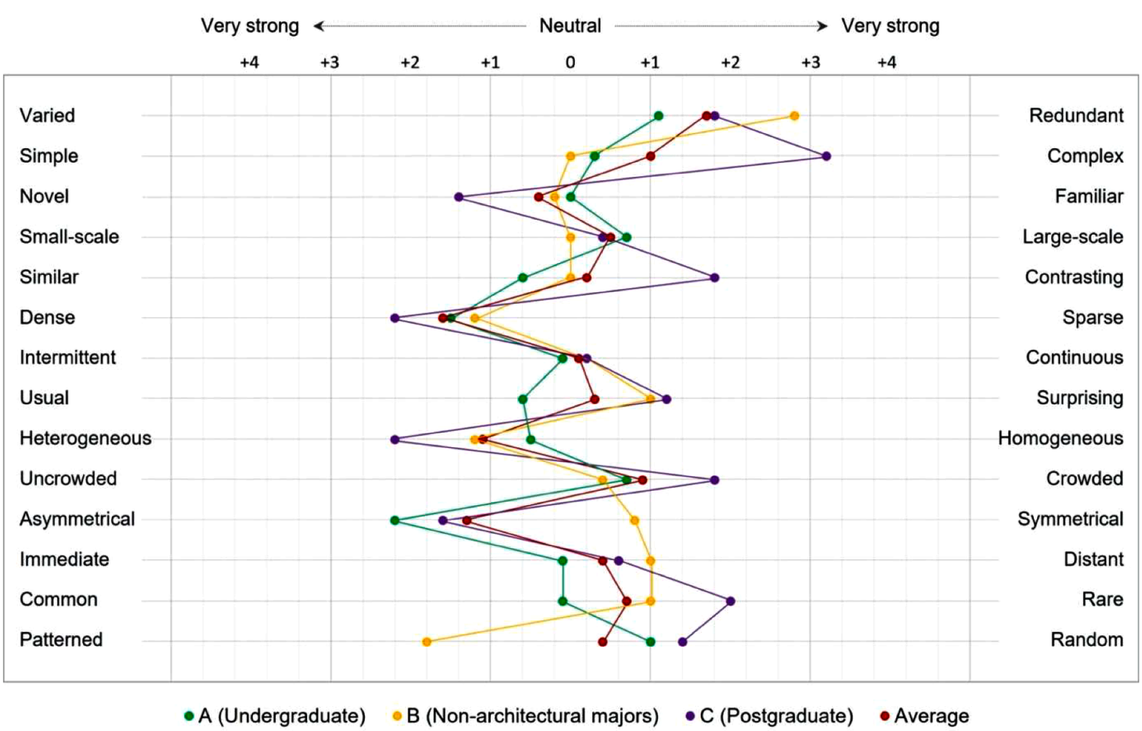


Fig. 15. Feelings for least preferred image (using information rate).

the space," "I feel friendly," "The space gave me a feeling of ease," and "I wasn't nervous" were notable. Examining each group with a relatively similar distribution, the postgraduate student group (C) awarded higher scores than the other groups, with responses of "I want to explore space," "I prefer more complex space structures," and "I wish there were people." While the graphs of the undergraduate (A) and postgraduate students (C) were similar, the non-architectural major group (B) was the most passive in space exploration and preferred a simpler form than the currently given space, as well as an environment without people.

The subjective opinions on the metaverse experienced included "It is virtual reality, but it has a lot of details, so it feels more realistic than I thought" and "Looking at the environment (the world) through tools, I felt that my gaze was limited." In addition, they responded that they wanted to add various behavioral functions such as sitting on a sofa, jumping, opening a window, picking up an object, and having a conversation. In the opinion of the non-architectural group (B), if the metaverse had multimedia elements, such as dialog, sound, and characters, a more realistic experience would be possible. However, the architecture majors (A, C) imagined the type of floor plan as a whole during their exploration and wanted a blueprint showing their current location. They also questioned the parts that made a building feel unrealistic, such as the fact that the house had windows on the roof, but it could not be climbed because there were no stairs, or an error in which the wall was pierced by object collision.

Regarding territoriality, a question on the arrangement of important objects and the space in which a participant wanted to stay was asked. The classification of important objects can vary from person to person, and because objects generally tend to belong to space, special definitions were provided to the objects in the experiments. Each subject was presented with a magic wand capable of doing anything. It was assumed that this cane belonged to the test subject and was an important object that only the subject could use. The subject was asked to indicate where he or she would like to place the wand in the main house and to explain why they choose that place. This question was not numerically relevant, because it only obtained the individual data necessary to generate the images to be provided together in the second survey. Subjective opinions on this question could synthesize opinions within the scopes of a place that was convenient to carry or use, a private place, a safe and deep place, place where the stay time was expected to be the longest, and a place where a remote control was typically placed.

Similarly, numerically synthesizing the question “Where would you like to stay if you spent time alone” was not statistically significant. Most subjective opinions on this topic were that it was cozy, comfortable, quiet, warm, and a place to rest. Other opinions included that it was good to provide privacy, a space to work or for activities, good concentration of the mind, nice sceneries, and enough space to invite people.

- Factor 3: Opinions on image sets

Factor 3 consisted of questions about the overall feelings and thoughts regarding the six image sets. Inquiries were made about how the participants felt about all of the derived images, the most or least preferred image set, and the reason for the selection. Mehrabian and Russell’s information rate was used to determine the impressions of the images.

Fig. 11 plots the response results for how the participants felt about all six image sets. Based on the averages of the opinions (red graph), the responses were close to the directions of “varied,” “complex,” “novel,” “large-scale,” “similar,” “dense,” “continuous,” “surprising,” “homogeneous,” “distant,” and “patterned.” In the responses of the non-architectural major group (B), “complex,” “novel,” “large-scale,” “dense,” “distant,” and “rare” recorded relatively high values. This Group B previously had a majority of neutral opinions on the metaverse, was relatively passive compared with the other groups in exploration, and preferred a simpler spatial structure in the experience of the metaverse. Even in the opinions on the image sets, Group B appeared to perceive the images as complex, large, and densely unfamiliar. The differences between the undergraduate (A) and postgraduate (C) groups were found with respect to the “heterogeneous–homogeneous,” “asymmetrical–symmetrical,” and “patterned–random” items. Group A perceived the images as symmetrical, patterned, and homogeneous, whereas Group B perceived them as asymmetrical, random, and heterogeneous.

Fig. 12 plots the selections of the most preferred image sets. The most preferred set followed the order No. 2, No. 4, and No. 5. Set 2 was a high-quality photo format made by Dall•E and represented the interior and exterior images of a building at a similar level to photography. The reasons for the preference were familiarity and good visibility, realistic appearance, clean and spacious view, and ease of space perception with photorealistic rendering. Set 4 was a floor plan created with Midjourney, and the subjects indicated that “it was the best format with which to understand the space,” “it was complex and interesting while giving a sense of order,” “the drawing was good, and it seemed to be in the game.” Set 5 was also a 3D-rendered format created by Midjourney. The reason it was preferred was as follows: “its format was good for understanding the space at a glance,” “I can imagine using it for seeing the inside and outside of the building,” “it felt familiar as an architectural digital modeling image,” and “it showed various results with applied expression techniques.” The most preferred image sets for Groups A–C were sets 4, 2, and 5, respectively, showing their slightly different preferences.

Fig. 13 plots the least preferred image sets. Overall, the opinions seemed to converge for a specific set of images. Among the least preferred image sets selected by all the subjects, Set 1 was the most common. Set 1 was a colored floor plan created by Dall•E. The major opinions were “dizzy and complicated,” “difficult to see and distinguish properly if you do not know the floor plan,” “difficult to understand what kind of drawing it is for non-majors due to poor readability,” “drawing elements and the way of expression present difficulty in recognizing space,” and “does not reflect the actual house.” The second non-preferred image set was set 3, which was a surrealist image created by Dall•E and applied in the Vladimir Kush style. Subjects answered that the reason for their dislike was “complicated, unrealistic appearance,” “interesting picture but the feeling of seeing it in a fairy tale or cartoons,” and “expressing a building but violating the laws of physics.” For the third non-preferred image set, the participants selected Set 6. Set 6 was a surrealist image created with Midjourney. The opinions included “It feels more like a picture than a house, it doesn’t give a sense of space” and “it is dizzy and difficult to understand.”

- Factor 4: Opinions on individual images

Factor 4 entailed selecting the most preferred and least preferred single images and responding with feelings and opinions about them. The questions of Factor 3 were meaningful for figuring out preferences or non-preferences for image formats by numerically synthesizing the results. As for Factor 4 though, the given 24 images were different for each individual; therefore, it was irrelevant to synthesize which specific image was selected. However, aggregating impressions of preferred and non-preferred images could be helpful in identifying the overall perceptions of preferred and non-preferred images.

Fig. 14 plots the information rates for the most preferred image. The average value of the subjects was characterized by items including “varied,” “dense,” “continuous,” and “uncrowded.” In particular, the “varied” and “dense” items showed similar values for all three groups. It could be inferred, then, that the subjects were slightly more interested in colorful images and high-density expressions. Based on each group, the results of the non-architectural major (B) and postgraduate (C) groups for the “large–small scale,” “intermittent–continuous,” and “uncrowded–crowded” items differed to opposite extremes. The non-architecture majors (B) answered that their preferred image was a harmonious one that was not complicated or large in scale. The postgraduate (C) group showed high values for “similar,” “usual,” “homogeneous,” “patterned,” “symmetrical,” and “dense” items. Group C mentioned architectural modeling, floor plans, and physical laws in their subjective responses. In connection with this result, it was inferred that when the postgraduate (C) group saw a building or an architectural image, they substituted an expression method of a building they had seen and learned about previously. The undergraduate group (A) showed the greatest convergence to the average graph. They selected various new, asymmetrical, and continuous images in the items above as their “slight preference.”

Subjective common opinions on the preferred images were “clean,” “variety, diverse,” “openness, wide view,” “easy to understand structure,” “easy to understand,” “interior and exterior can be seen at once,” “3D feeling,” “stability and warmth,” “feeling to live in,” “well-matched colors," and "harmony."

Fig. 15 plots the results for the non-preferred image. The average values of the subjects stood out in terms of "redundant," "complex," "dense," "heterogeneous," and "asymmetrical." Specifically, it can be assumed that the subjects had negative opinions about the images that were difficult to understand and mixed with heterogeneous things that felt complex and unnecessary. Based on each group, similarly to the preferred images, there were items with a wide range of differences in both the non-architectural major (B) and postgraduate (C) groups. Their opinions on differences spanned "simple–complex," "similar–contrasting," "asymmetrical–symmetrical," and "patterned–random." The non-architectural major (B) group responded that they felt that the images that they did not prefer were redundant, simple, patterned, and symmetrical. However, the postgraduate (C) students responded that their unpreferred images were generally complex, dense, heterogeneous, and random, and evocative of new feelings.

The responses of the two groups in fact were slightly contradictory. The postgraduate (C) group felt uncomfortable with complex images that were difficult to understand and far from reality, whereas the non-architectural majors (B) tended not to prefer images that were easy to read. The undergraduate group (A) converged to the average, similarly to the feelings for the preferred image. This group showed the highest values for "similar" and "asymmetrical" items.

The images they did not prefer were "too complicated and dizzy," "difficult to grasp structure, arrangement, and spatial composition," "heterogeneous," "would be inconvenient to use," and "too colorful and the combination was not good," "living room, wall, ceiling, etc. are unrealistic," and "symmetrical."

In the last part of the second questionnaire, the subjects were informed that the provided images were based on their spatial experience in the metaverse. And their opinions were solicited about what it would be like if a building like the image was automatically created in a virtual space. Eleven of the twenty subjects responded positively. The positive opinions included "It seems to be fun and interesting," "curiosity to experience through the metaverse," "new and diverse experiences," and "utilization would be high in the design process," "think I can quickly create a metaverse composed of wider and more colorful buildings," and "I want to decorate the space I want to live in." Seven participants provided neutral opinions. They mentioned that "it would be nice to be able to do some activities created in a virtual space, but I don't want to stay for a long time like in real life," "it is nice, but it seems scary and lengthy," and "the work has extreme likes and dislikes," "it will be possible to use it seamlessly if it becomes a space without major distortion compared to reality." The following two negative opinions were expressed: "It is dizzy because it is hard to see, and I do not think I will particularly like it," and "I do not think it is new because I think it is already common in virtual spaces."

6. Discussion

6.1. Visual information and preferred response to architectural space

This study assumed that in the future, AI could maximize efficiency by automatically creating user-customized spaces instead of modeling buildings individually in a metaverse environment. Under these assumptions, we analyzed the perceptions and opinions of how an architectural space image generated using user activity data feels to a user, who is the data contributor, through experiments.

To generate an architectural image according to user activity, it is necessary to first observe and interpret the spatial behavior of the user, capture the characteristics, and reflect them in the image creation. To this end, the literature review examined theories on the relationship between a built environment and human behavior. Among environmental psychology theories, the preference framework of Kaplan and Kaplan (1989) and the approach–avoidance theory of Mehrabian and Russell (1974a) study the preferences and emotional responses of people to an environment. We referred to these studies to establish the theoretical framework for our analyses.

According to the research by Kaplan and Kaplan, information is the key factor for the survival of organisms, and human preference responses appear to be affected by functional aspects, such as obtaining information. They suggested that considering the landscape from an informational perspective would help discover what underlies these preferences. This viewpoint can be linked to the analysis of the experimental results, which are presented in Section 5.2. From the subjective answers linked to the preferred image sets (Fig. 12) and image (Fig. 14), the participants indicated that they preferred easy-to-understand, familiar, and readable images. Moreover, they mentioned that "it is good to understand at a glance" and "I could imagine using it because I could see the exterior and interior of the building." Contrastingly, in the subjective answers linked to the non-preferred image sets (Fig. 13) and image (Fig. 15), the negative responses were complexity, difficult to distinguish, poorly readable, and poorly expressed drawings in the formal language. Thus, the subjects showed a preference for environments and architectural spaces that were easy to understand in terms of information about the environment, in agreement with Kaplan and Kaplan, and a non-preference for the opposite. This result suggests that it is possible to create a virtual building with a high probability of preference in the future if ambiguous or missing information is supplemented beyond the relationship between preference and non-preference. In this study, the buildings in the virtual space were limited to architectural images. For a text–image generator, the reflection of such environmental information is expected to produce slightly better results if it is ordered through a clear query at a low-level approach. However, fundamentally, much learning about architectural images is needed, and complete information can be reflected only when it evolves into a more intelligent deep learning model and enables semantic understanding.

6.2. Differences in preference responses according to degree of architectural education

In the present research, the participants were divided into three groups: undergraduate, non-architectural, and postgraduate students. The groups were distinguished on the scale of knowledge of architectural planning and design and familiarity with the relevant languages, forms, and tools of expression. The non-architectural majors had received education beyond university or graduate school but not architecture-related education. The undergraduate students had completed an average of (1 to 2) y, and the postgraduate students had received an average of (8 to 10) y of education and related research. Based on the experimental results from this viewpoint, there were slight differences in their perceptions of the metaverse, their experience of the virtual space for the experiments, and their opinions on the images.

First, the non-architectural majors assumed a neutral position in using the metaverse. In terms of experiencing the virtual space, compared with the other groups, they were more passive in space exploration and preferred a simpler spatial structure. In terms of opinions on images, "complex," "novel," "large-scale," "dense," "distant," and "rare" items scored relatively high. The non-architectural majors showed the greatest contrast with the postgraduate student group. The postgraduate group provided the most positive responses among the three groups regarding the use of the metaverse. For the metaverse space experienced in the experiments, there was a tendency to prefer space exploration and spatial structures that were more complex. In image recognition, higher scores were given to "continuous," "homogeneous," "symmetrical," and "patterned" items relative to the other groups.

Comparing the non-architectural majors with the postgraduate group, the former tended to regard the virtual space as new, strange, and complicated. However, the postgraduate student group was active in space exploration and showed a proclivity to familiarize and synthesize unfamiliar images as homogeneous, continuous, and patterned. The two groups also showed differences in their viewpoints on the subjective answers. The non-architectural majors group explained with expressions such as "a space I want to live in," "ideal," and "colorfully decorated." However, the postgraduate student group described "space elements," "finishing materials, furniture and interior props," "identification of structure," "expression technique," "physical laws," and "user point of view" from the viewpoint of constructing and evaluating buildings.

The undergraduate group showed the most similar trend in the average values of the three groups for all items. However, some contrasting trends were observed in the postgraduate student group in terms of the overall feeling for the images and the preference/non-preference for individual images. In the opinions on the entire image set (Fig. 11), the undergraduate and postgraduate student groups differed on the "heterogeneous-homogeneous," "asymmetrical-symmetrical," and "patterned-random" items. In the feelings for the preferred image (Fig. 14), there was a tendency to diverge to both extremes: "asymmetrical-symmetrical" and "patterned-random" items. This was also true in the feeling for the non-preferred image (Fig. 15): "simple-complex," "similar-contrasting," "asymmetrical-symmetrical," and "patterned-random" items.

The three groups exhibited slightly different tendencies in their preferred image formats as well (Fig. 12). The non-architectural majors preferred the high-quality photo format created by Dall•E, the undergraduates the floor plan format created by Midjourney, and the postgraduates the 3D-rendered format image set created by Midjourney.

Overall, the group that had long explored architectural images and engaged in spatial exploration, namely the postgraduate student group, had a strong tendency to explore spaces with an open mind. When evaluating architectural images, they seemed to apply a framework of logic, knowledge, and expression techniques to their previous architectures. It can be inferred that when they saw an architectural image, they interpreted the space and converted it into three dimensions. Therefore, they did not regard an image simply as a picture, but had a view of whether it could be built. However, the non-architectural majors felt that the virtual space was unfamiliar and complex and tended to perceive it as a complex, new, large-scale, and unfamiliar image in the image evaluation. In addition, the majority of this group tended to prefer photographic images to 2D floor plans or abstract images that required imagining unfamiliar spaces, and seemed to follow a more intuitive style in imagining spaces.

6.3. Comparison with previous studies using text-to-image generator

Since this study used a specially designed system to achieve the proposed purpose, studies with a similar structure and methodology are rare. However, it was possible to compare the methods and results with previous studies in terms of utilization and exploration of the text-to-image generator (hereafter referred to as the generator), the organizational methods of the queries, experiments, and interviews, and the models used. Seneviratne et al. (2022) and Paananen et al. (2023) were judged to be suitable for comparison in terms of research methods, results, and utilization of generative models (Table 12).

The items for comparison were the research goal and methodology, the use of the generator as a major tool, and the research results. The goal of this study was to create an image containing the user's active characteristics and examine the user's reaction to it. On the other hand, Seneviratne et al. (2022) investigated the quality of urban design and sought to utilize generators in urban planning. Paananen et al. (2023) tried to use the generator as a tool to support creativity. Seneviratne et al. (2022) were limited to exploring the possibility of using a tool called the generator, whereas this study and Paananen et al. (2023) used the generator as a tool, but investigated users' reactions to determine the potential efficacy and the possibility for popular use in the future.

There were also differences in the core methodology depending on the research goal. In this study, a theoretical framework and metaverse environment were established, and experiments and surveys were conducted. Seneviratne et al. (2022) constructed a grammar for query generation, and generated and evaluated images accordingly. Paananen

Table 12
Comparison with prior studies, research methods and results, and utilization of text-to-image generators.

Item		This study	Seneviratne et al. (2022)	Paananen et al. (2023)
Research goal		• Propose automatic generation of architectural space according to unconscious user activities in the metaverse, and analyze users' perceptions and opinions about it	• Exploring the potential of text-image generators to investigate urban design quality by generating image datasets relevant to urban design across multiple formats and regions.	• Exploring the potential of text-image generators to support creativity during the early stages of the architectural design process
Core research methodology		• Build a theoretical framework and metaverse environment for behavioral analysis • User participation in the metaverse • Survey on generated images	• Suggest how to construct a grammar for query generation • Generate queries related to urban environments • Evaluate the generated image	• Give students short-form architectural design tasks and text-image generators • Surveys, group interviews, and prompt analysis
Utilization of text-to-image generator	Utilization goal	Check the possibility of generating architectural images of text-to-image generator	To assist human experts in the built environment and urban planning	Creativity supports students' development of architectural design concepts
	Design stage	Planning and design phase	Planning and design phase	Early design phase
	Model	Dall·E 2, Midjourney	Dall·E 2	Dall·E 2, Midjourney, Stable Diffusion
	Parameters for generating text and images	• Let users experience the metaverse space, and collect user activity data in the space. • Analyze collected activity data for time, frequency, territoriality, exploration, and preferences for simple/complex space use. • Configure queries to generate architectural images by reflecting characteristics of a space considered to be preferred	• A query is created by combining the four relevant dimensions of urban design (image format (F), geography (G), urban quality (Q), and urban unit (U)) according to ordering rules. • Sequence rules for expanding the term: image format → geography → quality → unit	• Design subject (culture center) and site are given in common • Encourage students to express their concepts visually, including floorplan, interior perspective visualization, and facade material samples
Results		• Users responded positively when creating a virtual architectural space based on their own data. • Image creation based on user data can be an interesting approach, revealing the possibilities of AI architects • It was found that there is a difference in response to image preference according to architectural education training.	• Determined that the text-image model is powerful for generation of realistic images across different domains • Can accelerate the design process, including graphic design • Discovered potential to help human experts in this field	• Generation tools enriched the design process by supporting the serendipitous discovery of ideas and creative thinking. • Architects and designers could promote effective communication by using generators in the design process and in training

et al. (2023) conducted an experiment in the form of giving students tasks, and analyzed questionnaires, interviews, and used prompts.

The generators used as core technologies in the three studies could be compared in terms of their utilization goals, target design stages, models used, and parameters for text and image generation. In this study, the main utilization goal of the generator was the ability to express architectural space (e.g., a living room with many windows, a study with yellow walls, etc.). Seneviratne et al. (2022) intended to derive a vocabulary expressing the city and environment (e.g., health, safety, Melbourne, neighborhood, etc.). Paananen et al. (2023) observed the use of expanded models based on students' choices, focusing on the expression of architecture-related vocabulary (e.g., floorplan, interior perspective visualization, facade material).

In the three studies, the design stage using the generator was similar to the planning and design stage. The commonly used generator was Dall·E, and in addition, Midjourney and Stable Diffusion were utilized.

In the parameters set for the utilization of the generator, Seneviratne et al. (2022), in common with the present study, set systematic grammatical rules for query generation. In the generator, it is important to configure the input text specifically in order to output an appropriate image. Seneviratne et al. (2022) and the present study established combination rules and an order for the input text composition method. Paananen et al. (2023) also presented core vocabulary to students for use in construction of input text. Both Paananen et al. (2023) and Seneviratne et al. (2022) combined the input text in the system, though Paananen et al. (2023) differed in that students created various variations around the core vocabulary.

As for results, the present study showed that the virtual space creation method employed can be effective, as many participants responded positively to the architectural space created using their data. In addition, the analysis of preferred images and feelings about images using a comparison group showed that there were differences according to architectural education and training. Seneviratne et al. (2022) did not compare several models, but rather focused on Dall·E to determine the applicability of the model. Seneviratne et al. (2022) assessed that Dall·E is suitable for creating realistic images related to urban design and has the potential to help human experts in the field of urban planning. Paananen et al. (2023) showed that generators can enrich the design process by supporting students' creative thinking. Overall, they proved that generators can be effective tools in the design process, for education, and for communication.

All three studies explored the utilization of generators for specific tasks, and showed that they could be effectively utilized in the planning stages for architecture and urban planning. Although the currently used models have limitations in terms of image-generating performance and architectural representation, the three studies concluded that future models can produce more advanced results if they add data and training for related fields.

6.4. Limitations and implications

6.4.1. Limitations

This section deals with the limitations of the study and the problems that occurred during the experimental and development processes. First, the limitations of the research method and results are as follows. The experiments conducted in this study were focused on eliciting the characteristics and reactions of each participant. In order to obtain the results of the analysis of individual preferences, the participants in the study had to respond to two surveys after participating in the metaverse experience. Researchers also had to track the individuals participating in the experiment and perform individual analyses. The method and experimentation of this study obtained meaningful results, but were configured such that targeting of a large population was difficult. Therefore, in terms of results, it was not possible to produce statistics for a large population, and generalization of the results was difficult as well.

In terms of the limitations of the experiment and development process, both of the image generators (Dall•E, Midjourney) used in the experiments were sophisticated and large models that were difficult for individuals to learn. Also, since they are not specialized for architecture, they showed clear limitations in the processes of recognizing and creating space. The models were synthesized based on images that were in the training data. Therefore, the floor plan was created by synthesizing images to look like a floor plan, not a logical floor plan with meaningful content. These limitations can also be found in other studies using image generators in the field of architecture. Ploennigs and Berger (2023) pointed out that complex use cases such as floor plans or 3D model creation require much more specific training and situational understanding of the model part. Paananen et al. (2023) also concluded that their image generator failed to generate conventional floor plans, and, judging by the abstract representation, the training dataset for the floor plan images was insufficient for labeling for architectural design purposes. Bolojan et al. (2022) said that it is necessary to provide a complementary AI model due to limitations in language expression, because the learning of image generation models is based on large-scale general datasets that cannot capture the complex nuances of architectural design. In this regard, if, in the future, the learning data of the image generation model is supplemented, and the performance of the model is improved so that the data can be expressed more logically through semantic understanding of the building, it will be possible to create a higher-level architectural floor plan.

The metaverse space used in this study was designed based on the behavioral analysis framework with a clear purpose for the present experiment. The house in the virtual space was planned, from the composition and shape of rooms to the number of windows and lights. This controlled environment offered a utility for comparing experimental results and making conclusions, due also to the small number of variables. However, there are some differences from the general metaverse environment currently commercialized. The main difference is the interaction with other participants. In this study, each participant was allowed to enter and experience the virtual environment alone, but in a commercialized metaverse platform, a large number of people simultaneously access and interact with each other. As the interaction function in the metaverse gradually develops, interaction with others within the metaverse will become a more important factor. If we create an environment for interaction with others beyond what was planned in this study and accept sensory stimuli, we will be able to investigate richer opinions and feelings.

6.4.2. Implications

In this study, AI was applied to architectural planning, and a system architecture that can generate personalized images through analysis of user behavior data, rather than relying only on accidental image generation by AI, was devised. In addition, by investigating and analyzing how images created in this way feel to data providers, it was shown that a system that utilizes user data can elicit positive responses from users.

In order to utilize user data, data measurement and processing methods need to be established. In this study, grammatical rules by which to convert people's unconscious space experience data into text for creation of architectural space images was established. The proposed rules were systematically formulated based on traditional human-environment interaction theories such as environmental psychology. The rules presented in this study reflect a behavior analysis methodology for one virtual space, and although not absolute, they can be referred to in related studies.

Through a series of experiments, this study revealed the types of architecture people prefer in a virtual environment. As mentioned with regard to the theoretical background of this study (i.e., in the Literature Review), buildings in the virtual environment have the advantage of being free to build, because they are not subject to physical restrictions. Many virtual reality games that people enjoy have backgrounds, concepts, and rules that are not identical to reality. For this reason, we can assume that people may prefer architecture that is free from the constraints of reality. However, our research results indicated that people prefer buildings, and corresponding images, with forms similar to reality. In addition, this study analyzed preferences by dividing groups according to the degree of training in architectural education. The difference in response to images and the preferred image format in virtual reality were analyzed for each group, and it was found that the results were related to the degree of training in architectural education. These results can be referenced in the future when constructing buildings or developing environments in a virtual environment.

Finally, this study conducted an experiment using a deep-learning-based image generator and compared the methodology, model utilization, and research results with similar, previous studies. In the comparative analysis process, the image generator was used for certain goals and design stages for related studies, and the types of parameters used and the results of the final study were discussed. This data can be referenced as practical guidelines for future studies applying image generators in the field of architecture.

7. Conclusion

As a key industry for the future, the metaverse is growing in terms of people's interest in it and desire or willingness to use it. In that context, architecture in the metaverse is becoming a subject of interest as well. People want to continue their real-world activities in a metaverse, and thus, they require architectural spaces for rest, work, and play. However, it is quite inefficient to model cities and numerous buildings individually in a metaverse. In such cases, the use of AI can be maximized, and the building method can be diversified by reflecting interesting factors such as user-centered data processing or by creating results based on fortuity. Moreover, unlike in the real world, constructing a building in a virtual space is structurally free, as it is not subject to physical constraints such as gravity. Moreover, more creative attempts are possible too, because they are not constrained by either safety concerns or enormous construction costs.

In this context, the present study investigated how an architectural space automatically generated using user's unconscious activities in a metaverse environment would provide the user (i.e., the data provider) with certain feelings. To that end, we designed and produced a theoretical framework and metaverse space based on that framework to quantitatively analyze data activities with reference to the environmental psychology theory. We set grammatical rules by which floor plans and building images could be generated. Architectural space images were produced using a text-to-image generator based on deep learning. The derived images were shown again to the data provider, and their perceptions and opinions on the architectural space were collected in the form of their responses to a questionnaire. Although there were limitations to the representation of buildings depending on the learning data and the model's generation capabilities, users responded positively when virtual spaces were automatically created based on their data. Therefore, based on this study, such a user-data-based method will be an interesting approach in the future. Even for a metaverse, the goal should not be to build many buildings quickly but rather, to build buildings according to the preferences and interests of the users.

Future research can be a way to work on and advance the details of the proposed system. First, the theoretical framework for behavior analysis can be refined more elaborately if the results of this study are fed back. With the advancement of the metaverse experimental space, this task, as well as providing for multiple interactions in the metaverse (which lack was pointed out above as a study limitation) will be achievable. In addition, in future studies, we intend to approach generalization by supplementing the method of analysis and obtaining additional participants to secure more data. Furthermore, the results of this study will prove suggestive of a virtual-architectural-space design methodology that can identify design elements acting as key influencing factors. This methodology, we expect, can be derived by multi-faceted analysis.

CRediT authorship contribution statement

Sun-Young Jang: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Visualization. **Sung-Ah Kim:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential. But I have shared the link to my some possible data in this article.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1A2C1010759).

References

Abduljawad, M., Alsalmani, A., 2022. Towards creating exotic remote sensing datasets using image generating AI. In: Proceedings of the 2022 International Conference on Electrical and Computing Technologies and Applications (ICECTA). IEEE, pp. 84–88. https://doi.org/10.1109/ICECTA57148.2022.9990245.

Auliaramadani, M., Suciati, N., Hidayati, S.C., Fabroyir, H., Hariadi, R.R., 2020. Augmented reality for 3D house design visualization from floorplan image. In: Proceedings of the 2020 International Conference on Electrical Engineering and Informatics (ICELTICs). IEEE, pp. 1–6. https://doi.org/10.1109/ICELTICs50595.2020.9315422.

Barker, R.G., 1968. Ecological Psychology: Concepts and Methods for Studying the Environment of Human Behavior. Stanford University Press, Stanford, CA.

Bolojan, D., Vermisso, E., Yousif, S., 2022. Is language all we need? A query into architectural semantics using a multimodal generative workflow. In: Proceedings of the POST-CARBON, 27th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), 1, pp. 353–362. https://papers.cumincad.org/data/works/att/caadria2022_507.pdf.

Borji, A., 2022. How good are deep models in understanding the generated images? arXiv preprint. https://doi.org/10.48550/arXiv.2208.10760.

Canter, D., 1974. Psychology For Architects.

Chaillou, S., 2020. Archigan: artificial intelligence x architecture. In: Proceedings of the Architectural Intelligence: Selected Papers from the 1st International Conference on Computational Design and Robotic Fabrication (CDRF 2019). Springer Nature Singapore, Singapore, pp. 117–127. https://doi.org/10.1007/978-981-15-6568-7_8.

Clark, W.R., Ezell, J., Clark, J., Sheffield, D.N., 2009. Stay or leave: applying approach–avoidance theory to virtual environments. J. Database Mark. Cust. Strategy Manag. 16, 231–240. https://doi.org/10.1057/dbm.2009.25.

Croitoru, F.A., Hondru, V., Ionescu, R.T., Shah, M., 2023. Diffusion models in vision: a survey. IEEE Trans. Pattern Anal. Mach. Intell. 1–20. https://doi.org/10.1109/TPAMI.2023.3261988.

Davis, A., Murphy, J., Owens, D., Khazanchi, D., Zigurs, I., 2009. Avatars, people, and virtual worlds: foundations for research in metaverses. J. Assoc. Inf. Syst. 10 (2), 91–117. https://doi.org/10.17705/1jais.00183.

Demangeot, C., Broderick, A.J., 2010. Consumer perceptions of online shopping environments: a gestalt approach. Psychol. Mark. 27 (2), 117–140. https://doi.org/10.1002/mar.20323.

Dionisio, J.D.N., Burns, W.G., Gilbert, R., 2013. 3D virtual worlds and the metaverse: current status and future possibilities. ACM Comput. Surv. 45 (3), 1–38. https://doi.org/10.1145/2480741.2480751 (CSUR).

Elor, A., Song, A., 2020. isam: personalizing an artificial intelligence model for emotion with pleasure-arousal-dominance in immersive virtual reality. In: Proceedings of the 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020. IEEE, pp. 572–576. https://doi.org/10.1109/FG47880.2020.00091.

Evans, G.W., McCoy, J.M., 1998. When buildings don't work: the role of architecture in human health. J. Environ. Psychol. 18 (1), 85–94. https://doi.org/10.1006/jevp.1998.0089.

Fang, Y.W., Wu, C.H., Lee, F.P., Liu, W.M., 2012. Visitors' experiences of the art gallery at a teaching hospital. J. Exp. Clin. Med. 4 (3), 175–179. https://doi.org/10.1016/j.jecm.2012.04.009.

Fisher, J.D., Bell, P.A., Baum, A., 1984. Environmental Psychology. Holt, Rinehart, and Winston.

Gibson J.J., 1979. The ecological approach to visual perception. Boston: Houghton Mifflin.

Goldhagen, S.W., Gallo, A., 2017. Welcome to Your world: How the Built Environment Shapes Our Lives. Harper, New York.

Goodfellow I.J., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y., 2014. Generative adversarial nets. Advances in Neural Information Processing Systems 27 (NIPS 2014). https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf.

Huang, M.H., 2003. Modeling virtual exploratory and shopping dynamics: an environmental psychology approach. Inf. Manag. 41 (1), 39–47. https://doi.org/10.1016/S0378-7206(03)00024-7.

Kaplan, S., Preisner, W.F.E., 1973. Cognitive maps, human needs and the designed environment. In: Environmental Design Research, 1. Dowden, Hutchinson & Ross, Stroudsberg, PA, pp. 275–283.

Kaplan, S., 1975. An Informal Model For the Prediction of Preference. Hutchinson and Ross, Dowden.

Kaplan, R., 1985. The analysis of perception via preference: a strategy for studying how the environment is experienced. Landsc. Plan. 12 (2), 161–176. https://doi.org/10.1016/0304-3924(85)90058-9.

Kaplan, R., Kaplan, S., 1989. The Experience of nature: A psychological Perspective. Cambridge university press.

Kaplan, S., Kaplan, R., 2009. Creating a larger role for environmental psychology: the Reasonable Person Model as an integrative framework. J. Environ. Psychol. 29 (3), 329–339. https://doi.org/10.1016/j.jenvp.2008.10.005.

Kim, S.K., 2022. Metaverse II: Look ahead to the Future After 10 Years. Plan B Design. Korean.

Kulviwat, S., Bruner II,, G.C., Kumar, A., Nasco, S.A., Clark, T., 2007. Toward a unified theory of consumer acceptance technology. Psychol. Mark. 24 (12), 1059–1084. https://doi.org/10.1002/mar.20196.

Mehrabian, A., Russell, J.A., 1974a. An Approach to Environmental Psychology. the MIT Press.

Mehrabian, A., Russell, J.A., 1974b. A verbal measure of information rate for studies in environmental psychology. Environ. Behav. 6 (2), 233–252.

Munsell, A.H., 1905. A Color Notation. G. H. Ellis Co, Boston.

Næss, P., 2016. Built environment, causality and urban planning. Plan. Theory Pract. 17 (1), 52–71. https://doi.org/10.1080/14649357.2015.1127994.

Nauata, N., Hosseini, S., Chang, K.H., Chu, H., Cheng, C.Y., Furukawa, Y., 2021. House-gan++: generative adversarial layout refinement network towards intelligent computational agent for professional architects. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13632–13641.

Niland J., 2022a. BIG unveils designs for Vice's new virtual headquarters 'Viceverse'. Architect news. https://archinect.com/news/article/150300901/big-unveils-designs-for-vice-s-new-virtual-headquarters-viceverse (accessed 1 June 2023).

Niland J., 2022b. Zaha Hadid Architects is planning a metaverse city for Liberland. https://archinect.com/news/article/150300164/zaha-hadid-architects-is-planning-a-metaverse-city-for-liberland (accessed 1 June 2023).

Örer, Z., 2016. Emotional Responses Towards the Use of rectilinear, Symmetrical Curvilinear and Asymmetrical Curvilinear Forms in Virtual Environments. Bilkent Universitesi (Turkey) (Master Thesis,. http://hdl.handle.net/11693/32194.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., Lowe, R., 2022. Training language models to follow instructions with human feedback. Adv. Neural Inf. Process Syst. 35, 27730–27744.

Paananen V., Oppenlaender J. and Visuri A., 2023. Using text-to-image generation for architectural design ideation. arXiv preprint. 10.48550/arXiv.2304.10182.

Park, D., Kim, Y.J., 2019. Consumer behavior and perception for indoor rest place of department store corresponding to consumer characteristics-focusing on consumer cluster types in Seoul. J. Korea Inst. Spat. Des. 14 (7), 171–182. https://doi.org/10.35216/kisd.2019.14.7.171. Korean.

Park, S., Kim, H., 2021. 3DPlanNet: generating 3D models from 2D floor plan images using ensemble methods. Electronics 10 (22), 2729. https://doi.org/10.3390/electronics10222729 (Basel).

Park, S.M., Kim, Y.G., 2022. A metaverse: taxonomy, components, applications, and open challenges. IEEE Access 10, 4209–4251. https://doi.org/10.1109/ACCESS.2021.3140175.

Perkins, D.V., Burns, T.F., Perry, J.C., Nielsen, K.P., 1988. Behavior setting theory and community psychology: an analysis and critique. J. Community Psychol. 16 (4), 355–372. https://doi.org/10.1002/1520-6629(198810)16:4<355::AID-JCOP2290160402>3.0.CO;2-D.

Ploennigs, J., Berger, M., 2022. arXiv preprint. https://doi.org/10.48550/arXiv.2212.09399.

Ploennigs, J., Berger, M., 2023. Analysing the usage of AI art tools for architecture. In: Proceedings of the 2023 European Conference on Computing in Construction and the 40th International CIB W78 Conference. European Council on Computing in Construction. https://doi.org/10.35490/EC3.2023.253.

Porteous, J.D., 1977. Environment & behavior: Planning and Everyday Urban Life. Addison-Wesley, Reading, MA.

Proshansky, H.M., 1978. The city and self-identity. Environ. Behav. 10 (2), 147–169.

Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., 2021. Learning transferable visual models from natural language supervision. In: Proceedings of the 38th international conference on machine learning. PMLR 139, pp. 8748–8763. In: http://proceedings.mlr.press/v139/radford21a/radford21a.pdf.

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., Sutskever, I., 2021. Zero-shot text-to-image generation. In: Proceedings of the 38th international Conference on Machine Learning. PMLR 139, pp. 8821–8831. In: http://proceedings.mlr.press/v139/ramesh21a/ramesh21a.pdf.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., Chen, M. Hierarchical text-conditional image generation with clip latents. arXiv preprint 10.48550/arXiv.2204.06125.

Robert, D., John, R., 1982. Store atmosphere: an environmental psychology approach. J. Retail. 58 (1), 34–57.

Roose K., 2022. An A.I.-generated picture won an art prize. artists aren't happy. The New York times. https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html (accessed 1 June 2023).

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E.L., Ghasemipour, K., Gontijo Lopes, R., Karagol Ayan, B., Salimans, T., Ho, J., 2022. Photorealistic text-to-image diffusion models with deep language understanding. Adv. Neural Inf. Process Syst. 35, 36479–36494. In: https://proceedings.neurips.cc/paper_files/paper/2022/file/ec795aeadae0b7d230fa35cbaf04c041-Paper-Conference.pdf.

Schumacher, P., 2022. The metaverse as opportunity for architecture and society: design drivers, core competencies. Archit. Intell. 1 (1), 1–20. https://doi.org/10.1007/s44223-022-00010-z.

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

Sharma, D., Gupta, N., Chattopadhyay, C., Mehta, S., 2017. Daniel: a deep architecture for automatic analysis and retrieval of building floor plans. In: Proceedings of the 2017 14th IAPR international conference on document analysis and recognition (ICDAR), 1. IEEE, pp. 420–425. https://doi.org/10.1109/ICDAR.2017.76.

Smart J., Cascio J., Paffendorf J., 2007. Metaverse Roadmap Overview. https://www.w3.org/2008/WebVideo/Annotations/wiki/images/1/19/MetaverseRoadmapOverview.pdf (accessed 1 June 2023).

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S., 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In: Proceedings of the 32nd international conference on machine learning, 37. PMLR, pp. 2256–2265. In: http://proceedings.mlr.press/v37/sohl-dickstein15.pdf.

Song, S., Wang, Y., Liang, W., Li, X., 2022. Moving soon? Rearranging furniture using mixed reality. In: Proceedings of the 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops. VRW). IEEE, pp. 798–799. https://doi.org/10.1109/VRW55335.2022.00250.

Stephenson, N., 2008. Snow crash. Bantam Books, New York.

Wang, Y., Su, Z., Zhang, N., Xing, R., Liu, D., Luan, T.H., Shen, X., 2022. A survey on metaverse: fundamentals, security, and privacy. IEEE Commun. Surv. Tutor. https://doi.org/10.1109/COMST.2022.3202047.

Wise, J.M., 2000. Home: territory and identity. Cult. Stud. 14 (2), 295–310.

Witteveen S., Andrews M., 2022. Investigating prompt engineering in diffusion models. arXiv preprint. 10.48550/arXiv.2211.15462.

Wu, W., Fu, X.M., Tang, R., Wang, Y., Qi, Y.H., Liu, L., 2019. Data-driven interior plan generation for residential buildings. ACM Trans. Graph. 38 (6), 1–12. https://doi.org/10.1145/3355089.3356556 (TOG).

Yao C., 2022. Grimshaw and farshid moussavi create metaverse social spaces. https://www.dezeen.com/2022/12/16/grimshaw-hwkn-pax-world-metaverse (accessed 1 June 2023).

Yildirim, E., Hale Kozlu, H., 2022. Text-to-image generation ai in architecture. Art and Architecture: Theory, Practice and Experience. Livre de Lyon, pp. 97–120.

Yüncü, D., 2015. Relationships between perceptions of virtual destination environment, satisfaction and loyalty. J. Hosp. Tour. Technol. 6 (2), 160–173. https://doi.org/10.1108/JHTT-07-2014-0021.