ASME Journal of Mechanical Design
Online journal at:
https://asmedigitalcollection.asme.org/mechanicaldesign

An Artificial Intelligence-Generated Content-Enabled Personalized Design Approach for Proactive User Interaction in an Immersive Environment

Mengyang Ren
Department of Industrial and Systems Engineering,
The Hong Kong Polytechnic University,
Hong Kong 999077, China
e-mail: mengyang.ren@connect.polyu.hk

Junming Fan
Department of Industrial and Systems Engineering,
The Hong Kong Polytechnic University,
Hong Kong 999077, China
e-mail: jun-ming.fan@connect.polyu.hk

Wun Ting Chan
Laboratory for Artificial Intelligence in Design,
Hong Kong Science Park, New Territories,
Hong Kong 999077, China
e-mail: wtchan@aidlab.hk

Carman K. M. Lee
Department of Industrial and Systems Engineering,
The Hong Kong Polytechnic University,
Hong Kong 999077, China
e-mail: ckm.lee@polyu.edu.hk

Pai Zheng[1]
Department of Industrial and Systems Engineering,
The Hong Kong Polytechnic University,
Hong Kong 999077, China
e-mail: pai.zheng@polyu.edu.hk

Rapid advancement of artificial intelligence and immersive technologies is revolutionizing various sectors, notably product design. However, the traditional personalized design process, which depends on predefined elements with limited user input, often results in products that do not fully align with individual preferences and lack substantial user engagement. To fill this gap, the emergence of artificial intelligence (AI)-generated content (AIGC) presents a significant opportunity for mass personalization through natural language interactions. Inspired by this paradigm, this article proposes an AIGC-enabled personalized product design approach, which integrates a configuration retrieval model with a fine-tuned text-to-3D generative model (TAPS3D model), enabling users to create personalized products within an immersive environment. While the current system requires approximately two minutes for 3D shape generation, this level of responsiveness is considered suitable for concept exploration in early-stage design workflows, where rapid iteration is prioritized over instantaneous feedback. Furthermore, a case study is conducted focusing on the design of personalized steering wheels to demonstrate the feasibility of this methodology. Furthermore, the effectiveness of the proposed approach in improving user experience is evaluated using a comparative experiment with the traditional configuration system. The findings indicate that our proposed AIGC-enabled personalized design system effectively enhances personalization, facilitates user engagement, improves the interaction experience, and increases user satisfaction. [DOI: 10.1115/1.4069689]

Keywords: conceptual design, creativity and concept generation, generative design, product design, user-centered design

1 Introduction

In today's rapidly evolving market, manufacturers are compelled to adapt to increasingly diverse and specific consumer demands [1,2]. Traditional product design methods often fail to meet these requirements, thereby facilitating the progress of personalized design approaches. Personalized design seeks to tailor products to individual needs and preferences, fostering a more intimate alignment between products and user expectations [3]. This shift underscores the importance of value cocreation, where consumers are actively involved in the design process, enabling a closer match between product offerings and user requirements [4]. However, current design processes remain predominantly reactive in nature. Communication gaps between designers and consumers result in slow, labor-intensive iterative cycles [5]. Additionally, high entry barriers further restrict consumer participation in the design process. Most existing systems only offer limited personalization through predefined options, often lacking intuitive, multimodal, and immersive experiences for users [6,7].

Artificial intelligence (AI)-generated content (AIGC) offers a transformative solution that enables rapid, interactive, and personalized content creation [8–10]. AIGC shifts users from passive recipients to proactive cocreators [11,12]. Meanwhile, immersive technologies bridge physical and virtual worlds, improving realism, creativity, and user satisfaction [13,14]. Integrating AIGC with immersive environments can foster proactive user participation and enrich the design experience.

To address these challenges, this article proposes an AIGC-based personalized design approach within an immersive environment, supporting product generation and active user involvement. The feasibility of this approach is demonstrated through a case study on personalized steering wheel design, and its impact on usability and user experience is evaluated through an empirical study. The remainder of the article is organized as follows. Section 2 reviews related work, Sec. 3 introduces the methodology and system, Sec. 4 presents the case study, Sec. 5 reports the user evaluation, Sec. 6 discusses limitations and future work, and Sec. 7 concludes the article.

2 Literature Review

This section first reviews existing research on AI-based interaction design in immersive environments and text-to-3D mobile deployment and then summarizes the research gaps.

[1]Corresponding author.

Contributed by the Design Automation Committee of ASME for publication in the Journal of Mechanical Design. Manuscript received December 27, 2024; final manuscript received August 25, 2025; published online October 17, 2025. Assoc. Editor: Zhenghui Sha.

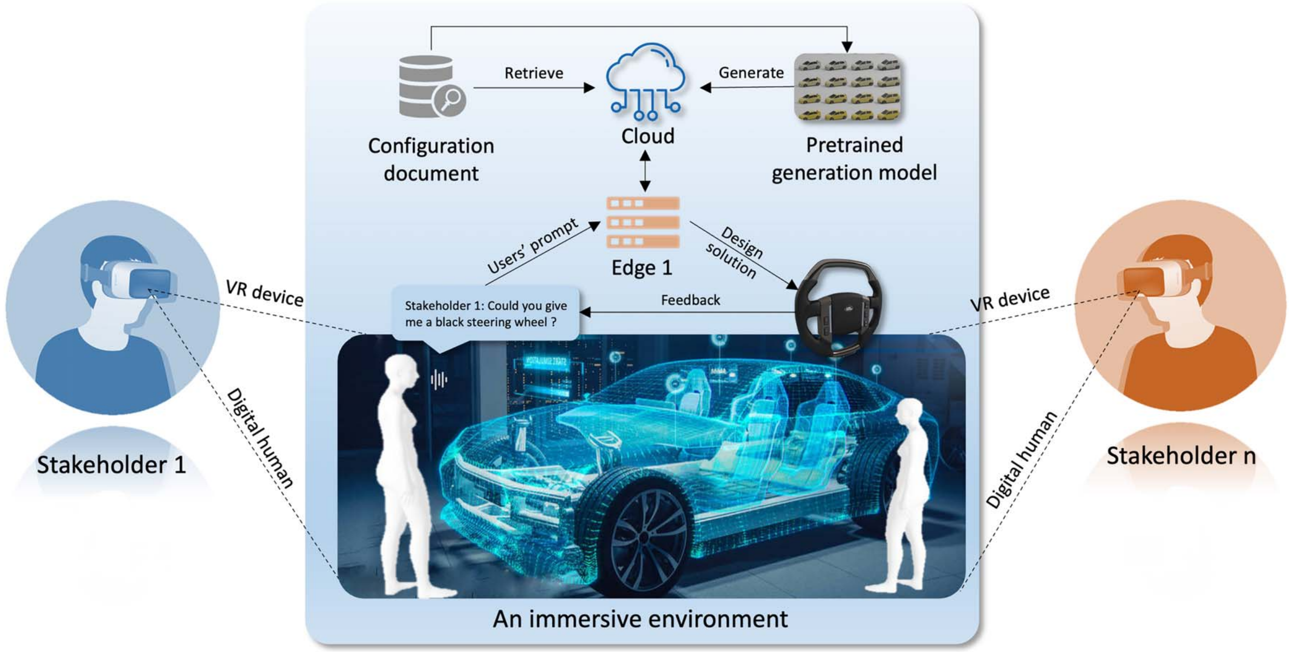

Fig. 1 The AIGC-enabled personalized design framework in an immersive environment

AI-based Interaction Design in Immersive Environments. The integration of generative AI models with immersive environments has gained attention in recent years [15,16]. For instance, Zhang et al. [17] proposed VRCopilot, providing an intuitive platform for users to create 3D layouts with pretrained generative models via direct manipulation in a VR environment. Rasch et al. [18] investigated AI representation modes for the 3D object cocreation process in VR settings. Stemasov et al. [19] proposed Mix and Match, an MR-based system that allows users to browse model repositories, preview models, and adapt them to the environment. Weng et al. [20] proposed Dream Mesh, a speech-to-3D model generation system in an MR environment, which leverages a stable diffusion model and instructs pix2pix to generate the 3D model based on voice input from the user. Behravan et al. [21] analyzed the current environmental information via vision language models and further leveraged the text-to-3D generative model to generate appropriate 3D objects in AR environments. Beharavan and Gračanin [22] present Matrix, a cutting-edge text-to-3D generative model in AR environments, which is equipped with a pregenerated object repository, further enables seamless user interactions through spoken commands. Xing et al. [23] present sMoRe, a mobile text-to-3D generative model that allows users to create and place virtual objects in mixed reality spaces through voice interaction while specifying spatial constraints.

Text-to-3D mobile deployment. The rapid progress of text-to-3D generative models made it possible to deploy on mobile platforms. Most researchers adopt the image method as the foundation for text-guided 3D generation and deploy them on the phone. For instance, Truong and Le [24] proposed MobileGen3D, which utilizes multi-view real-world images and a textual instruction to produce high-fidelity 3D content that resembles the provided real-world content and demonstrates a case study about character generation. Liu et al. [25] presented DreamStone for text-guided 3D shape generation based on a pretrained single-view reconstruction model without paired text and 3D data. Li et al. [26] proposed LLM4CAD for engineering design, leveraging the capabilities of multimodal LLMs to generate a computer-aided design (CAD) model.

In summary, studies have explored various aspects of integrating generative AI with immersive environments and mobile platforms. However, these works typically focus on pretrained models and do not address the challenge of enhancing the performance of 3D generative models within a specific domain and ignore the current configuration elements.

3 Methodology

To fill the gap, this section proposes an AIGC-enabled personalized design approach and elaborates its system architecture.

3.1 AIGC-Enabled Personalized Design Framework.

The AIGC-enabled personalized design framework is illustrated in Fig. 1. Various stakeholders can enter the immersive design environment by wearing a VR head-mounted display (HMD) and articulating their personalized product design requirements to interact with the graphical user interfaces (GUIs) for virtual product design innovation. User voice inputs are processed locally on edge devices, which access cloud data remotely to retrieve corresponding 3D assets from the configuration document or generate suitable 3D objects using a pretrained generative model.

Additionally, the body avatarization is provided to users, enabling the creation of digital humans that serve as a link between real and virtual spaces. This allows for the accurate reflection of users' movements, enabling them to observe their actions in the virtual environment in real time and interact with the retrieved or generated models from a first-person perspective.

3.2 System Architecture.

The system architecture of the proposed framework is depicted in Fig. 2. Based on the workflow, this structure can be categorized into three distinct modules: the requirement identification module, the configuration retrieval module, and the pretrained generative model.

3.2.1 Requirement Identification Module. An immersive design environment is established utilizing VR HMDs and Unity software, allowing users to interact with product design innovations. User requirements are effectively gathered through voice recognition and natural language processing tools. Robust cloud-based speech recognition services, such as the DictationRecognizer application programming interface, Google Cloud Speech-to-Text,[2] and Microsoft Azure Speech Service,[3] can be employed within VR settings to facilitate real-time conversion of user voice input into text. Integration of these services in Unity is achieved via plugins or

[2]https://cloud.google.com/speech-to-text
[3]https://azure.microsoft.com/en-gb/services/cognitive-services/speech-to-text/

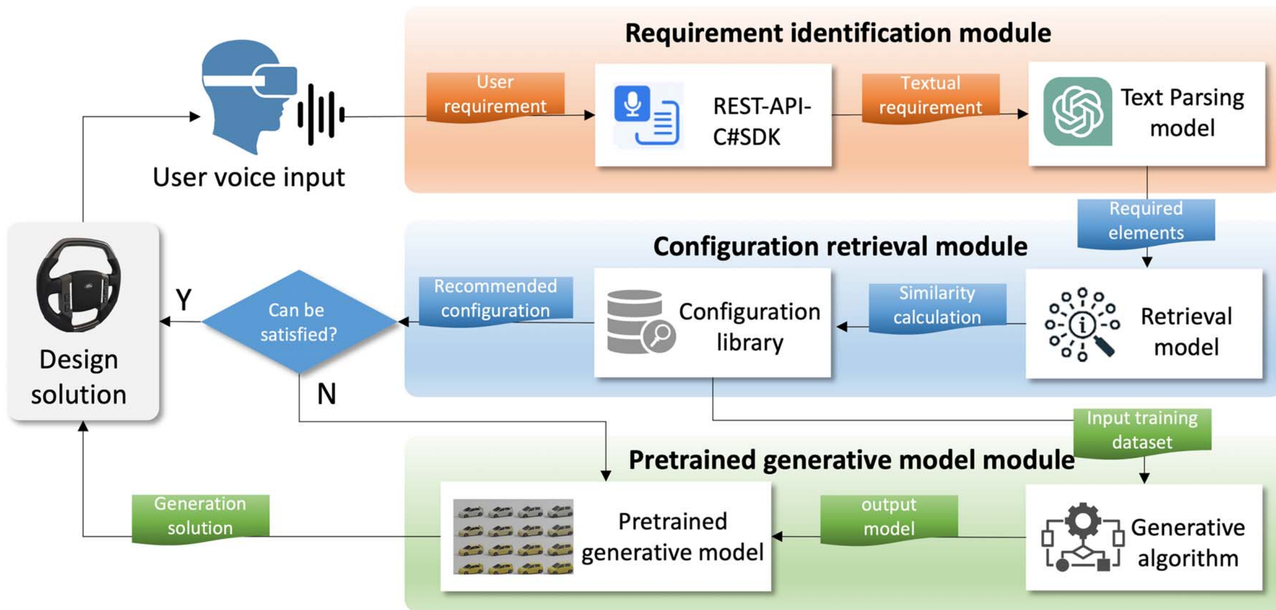

Fig. 2 The system architecture of the proposed AIGC-enabled personalized design framework

software development kits. Scripts are developed to capture voice input and transmit audio data to the cloud for processing. The converted text is then analyzed using a text parsing model to extract user intentions and needs. The results are used to trigger appropriate VR interactions.

3.2.2 Configuration Retrieval Module. In the configuration retrieval module, we employ Word2Vec [27] combined with cosine similarity to retrieve configuration elements, as illustrated in Fig. 3. The structure of the configuration document consists of a variety of parameters and descriptive information for different products. To enhance computational processing and understanding, these documents are converted into high-dimensional vector representations and stored in a vector database. Retrieval is performed by calculating the cosine similarity between the user requirement vector and each configuration vector, thereby identifying the most relevant configuration element. Cosine similarity (*Sim*) is defined as

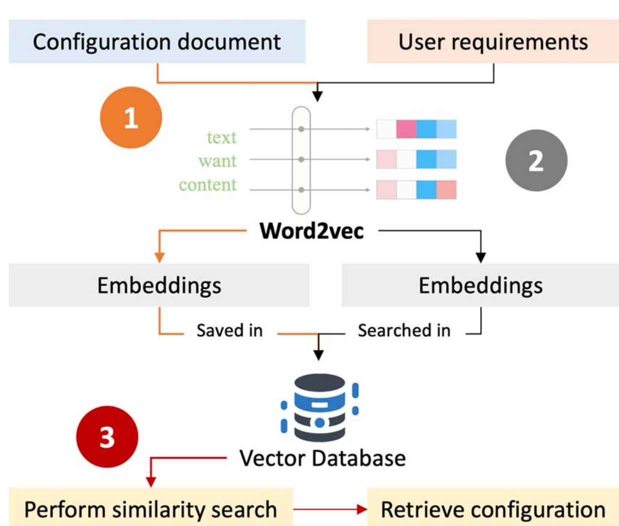

Fig. 3 The process of configuration retrieval

$$Sim\left(\vec{q}, \vec{d}\right) = \frac{\vec{q} \cdot \vec{d}}{\|\vec{q}\|\,\|\vec{d}\|} = \frac{\sum_{i=1}^{n} q_i d_i}{\sqrt{\sum_{i=1}^{n} q_i^2}\sqrt{\sum_{i=1}^{n} d_i^2}} \tag{1}$$

where $\vec{q}$ and $\vec{d}$ denote the query vector and the configuration document vector, respectively.

3.2.3 Pretrained Generative Model. In this subsection, we fine-tune the TAPS3D model [28] to enable personalized design generation. TAPS3D is a text-to-3D generative model capable of producing high-quality, textured 3D meshes with arbitrary topologies, guided by user-provided textual instructions. The model builds upon the principles of NVIDIA's GET3D [29] and leverages pseudo captions generated by the gpt-4o model. The overall workflow is illustrated in Fig. 4, which highlights three main components: pseudo caption generation, 3D-textured mesh generation, and authenticity evaluation.

To train the generative model, a dataset of 3D objects is rendered into multiview images. The multimodal large language model gpt-4o is used to generate descriptive pseudo captions for these images, reflecting how users typically express design requirements. The CLIP model [30] encodes these captions into embeddings, which serve as input for the 3D mesh generator.

The generator synthesizes 3D shapes and textures conditioned on the caption embeddings, producing high-fidelity, textured meshes. A differentiable renderer is then used to generate 2D images from the synthesized 3D models. To ensure the generated objects are realistic and consistent with the input captions, one image discriminator evaluates the authenticity of the rendered images. The training objective is to maximize the similarity between the generated images and the input captions, as measured by the CLIP model, and to encourage visual fidelity through an image regularization loss:

$$\mathcal{L}_{\text{clip}} = 1 - \cos\left(E_i(I_x), E_t(t)\right) \tag{2}$$

where $\mathcal{L}_{\text{clip}}$ denotes the loss value between the captions and the rendered images. E_i and E_t are the image and text encoders of the CLIP model, respectively. I_x refers to the rendered images, while t signifies the pseudo captions text. Considering the limited variety of captions, it is recommended that a low-level image regularization loss should be introduced to enhance the performance of the generative

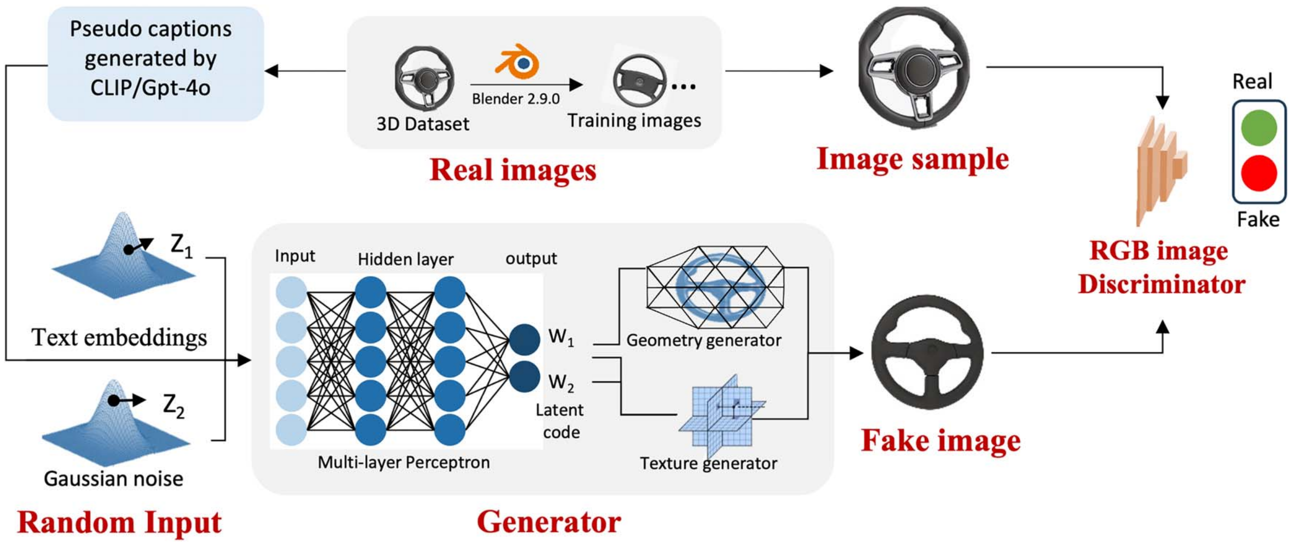

Fig. 4 Overview of the text-guided 3D generative model

objects, as represented by

$$\mathcal{L}_{\text{img}} = 1 - \cos\left(E_i(I_x), E_i(I_x^{gt})\right) \tag{3}$$

where $\mathcal{L}_{\text{img}}$ represents the low-level image regularization loss value, and I_x^{gt} refers to the ground truth rendered images. Therefore, the objective of model training of the generative model can be articulated as follows:

$$\mathcal{L} = \mathcal{L}_{\text{clip}} + \mathcal{L}_{\text{img}} \tag{4}$$

Furthermore, a smaller value of $\mathcal{L}$ indicates a better generation performance in alignment with the pseudo captions.

4 Case Study

In this section, a case study on personalized steering wheel design is conducted, aiming at illustrating the applicability of the proposed method in a specific context.

4.1 Fundamental Setup. This subsection outlines the experimental scenario and system communication architecture. In the experimental setup (Fig. 5), the system comprises a driving simulator, a VR HMD, two controllers, base stations, and two host computers. Detailed hardware specifications are provided in Table 1. Host PC 1 is responsible for managing virtual interaction, while Host PC 2 is dedicated to train the text-to-3D generative model. The primary software tools utilized are Unity and Visual Studio Code.

In the system communication architecture illustrated in Fig. 6, orange arrows signify commands and green arrows indicate feedback. The information flow begins with the VR HMD. When users wear the VR HMD and launch the program, they encounter the initial GUI, shown in Fig. 7(*a*). Here, users can utilize a VR controller to click "Start" to voice input for personalized requests and "End" to stop speech recognition, as shown in Fig. 7(*b*). This functionality is enabled by Unity's DictationRecognizer tool. Two host computers act as service agents: Host PC 1 achieves immersive interaction, while Host PC 2 operates the pretrained text-to-3D generative model, both working to deliver personalized product recommendations based on user prompts.

4.2 Implementation Details. The following two subsections provide comprehensive information regarding the proposal for personalized design of steering wheels, which encompasses the contents of the steering wheel configuration files, along with the implementation of the pretrained text-to-3D generative model for steering wheels. Initially, the personalized steering wheel design solution is extracted from existing configuration documents. If the system investigates that the user is not satisfied with the retrieved design solution, it will employ the text-to-3D generative model to produce additional personalized steering wheel design options.

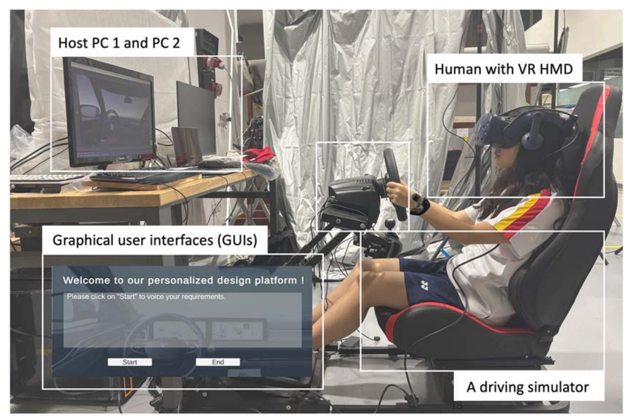

Fig. 5 The overall experiment scenario

Table 1 The system hardware setup

Hardware	Description
VR HMD	HTC VIVE Pro Eye with two controllers and base stations
Host PC 1	A laptop with an Intel Core i7-13700K CPU, one NVIDIA GeForce RTX 3090 GPU and Windows 11 operating system
Host PC 2	A laptop with an AMD Ryzen 9 7950X3D 16-Core processor, two NVIDIA GeForce RTX 4090 GPUs, Linux Ubuntu 20.04 operating system and CUDA version 12.3

4.2.1 The Retrieval of Steering Wheels. In this section, we present a configuration document that includes 10 elements randomly selected from the open dataset Sketchfab.[4] Each element is

[4]https://sketchfab.com/

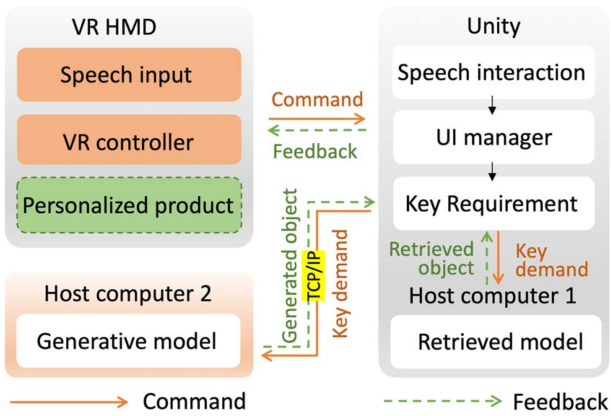

Fig. 6 The system communication architecture

annotated with an identifier (id), a detailed description, associated tags, and fileIdentifier. The purpose of this section is to introduce the structure of the configuration document, elaborating on the various parameters and descriptive information. Additionally, we demonstrate how user requirements are processed to calculate the similarity between the user's requirement and each configuration element. These elements are then ranked according to their similarity scores, allowing for the retrieval of the configuration element that exhibits the highest similarity.

4.2.2 The Generation of Steering Wheels. Database. A text-to-3D generative model is trained to achieve personalized steering wheel generation. We collected 280 3D steering wheel objects from open datasets like Shapenet [31], Objaverse [32], and 3D Warehouse.[5] After preprocessing the data using Blender 2.9.0, the dataset was split into training (70%), validation (10%), and test (20%) subsets. Duplicate test samples were removed. Each object was rendered from 100 random views due to the limited dataset size. During training, each rendered image was paired with a pseudo caption generated by gpt-4o, using the prompt: "Please describe the type, style, material, color, and components/feature of the object shown in this image. Your response should be a detailed description of the object without mentioning the background, in a single, short sentence of less than 10 words."

Training. Various text prompts produce a range of text-guided 3D objects. Textual features are concurrently processed by both the mesh and texture mapping networks. Notably, the semantics of the 2D-rendered images align well with the provided prompts. The generative model demonstrates robust capabilities in controlling both the geometry and texture of the steering wheels. All visualizations are generated from models that have been trained for 3000k iterations. For example, when the input prompt is "*a black racing steering wheel*," the generated samples precisely match the given text command. However, there is limited variability in the shape and texture of the resulting 3D objects from the same prompt.

Evaluation. We adopt the two loss functions, $\mathcal{L}_{\text{clip}}$ and $\mathcal{L}_{\text{img}}$, to quantitatively evaluate our text-to-3D generative model for the steering wheel. Specifically, $\mathcal{L}_{\text{clip}}$ facilitates the alignment learning between text and image representations, while $\mathcal{L}_{\text{img}}$ measures the similarity between the generated image and the real image. The minimization of these numerical values indicates a closer alignment between text–image and image–image representations. Therefore, we select the trained model checkpoint with the lowest loss values (1.54 for $\mathcal{L}_{\text{clip}}$ and 0.189 for $\mathcal{L}_{\text{img}}$) on the validation set for text-to-3D inference.

Results and comparison. Figure 8 shows diverse 3D visualizations of steering wheels generated by our trained text-to-3D generative model. To evaluate the performance of our model, we compare it with the state-of-the-art text-to-3D generative models: Meshy AI[6] and Tecent Hunyuan3D [33]. Qualitative results are presented in Fig. 8, demonstrating that our model generates high-quality 3D objects with diverse textures and shapes. Quantitative evaluation is conducted based on three aspects: semantic consistency, geometric fidelity, and texture fidelity. For each prompt, we generate 10 samples, and the evaluation results are detailed as follows:

(1) Semantic consistency comparison. We evaluate the average semantic similarity between input prompts and 3D objects generated using the CLIP score [34]. As shown in Table 2, our model achieves a relatively higher text-to-3D semantic consistency, indicating that the generated 3D objects align more closely with the input prompts. For Prompt 2, our model scores lower than Meshy AI, which may be attributed to the complexity of the prompt and limited training data. Overall, our model demonstrates good performance in generating personalized steering wheel designs that meet user requirements.

(2) 3D Geometry Quality. We assess the geometric fidelity of the generated objects by calculating the chamfer distance (CD) [35] between the generated point cloud and the ground truth point cloud. The average CD results are presented in Table 3. A one-way analysis of variance (ANOVA) reveals no significant difference among the models, suggesting comparable geometric performance.

(3) Texture fidelity evaluation. Texture fidelity is evaluated using the texture learned perceptual image patch similarity (LPIPS) score [36] computed between the generated texture UV maps and the ground truth texture UV maps. The average LPIPS scores for each prompt are reported in Table 4. A one-way ANOVA shows no significant differences among the models, indicating that our model performs on par with the other models in terms of texture quality.

In summary, these results indicate that our approach achieves improved semantic alignment while maintaining competitive performance in terms of geometric and texture quality.

4.3 Prototype System Implementation.

We implemented a prototype system through a specific personalized requirement, focusing on providing users with personalized steering wheel design solutions based on their specific requirements.

4.3.1 Interaction Process. Users wearing the HTC Vive Eye Pro device interact with the system via voice commands to express their preferences. They receive personalized steering wheel designs tailored to their specifications. Initially, the design solution is determined through a similarity matching process within the configuration document. If the user is not satisfied, the system employs a text-to-3D generative model. The user's voice input, such as "I want an antique, red racing steering wheel," is converted into a text requirement using the DictationRecognizer tool in Unity, as shown in Fig. 7(*b*).

In our system implementation, the steering wheel is automatically positioned within the immersive environment using a programmatic approach. This is accomplished by referencing predefined spatial coordinates and orientation data for the target location (e.g., the steering column), allowing the object to automatically "snap" into the correct position and orientation without requiring manual user intervention.

Regrading scaling adjustments of the generated 3D mesh, the system utilizes the interaction mechanisms provided by the SteamVR toolkit. Users can select and dynamically resize the generated models using both VR controllers through intuitive gestures such as pinch-and-drag or expand-to-scale. This interactive method allows flexible and precise control over object dimensions during the running process.

[5]https://3dwarehouse.sketchup.com/

[6]https://www.meshy.ai/discover

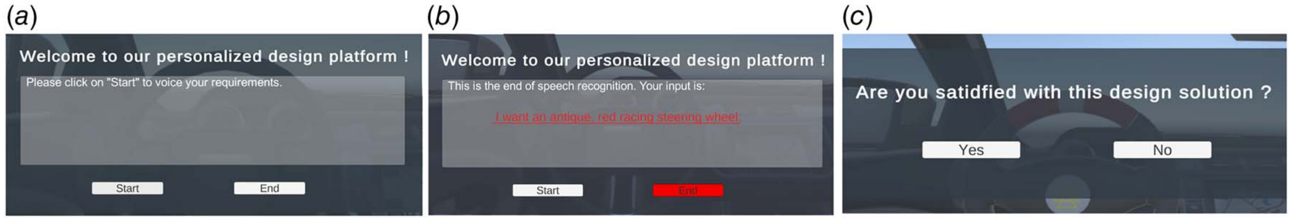

Fig. 7 The system interaction interfaces. (*a*) The initial interaction interface of this system. (*b*) The interface of speech recognition result. (*c*) The investigation interface of user satisfaction for the provided design solution.

Fig. 8 Comparison of three examples generated by the state-of-the-art text-to-3D generative models

Table 2 Semantic consistency comparison among three text-to-3D generative models using CLIP score [34]

Text prompt	Ours	Meshy AI	Hunyuan3D
Prompt 1	0.49	0.24	0.27
Prompt 2	0.41	0.42	0.54
Prompt 3	0.55	0.21	0.44

Table 3 Average CD [35] calculation for generated geometric fidelity evaluation

Type	Ours	Meshy AI	Hunyuan3D
CD_Prompt 1	12.98	13.63	15.03
CD_Prompt 2	2.05	1.08	1.70
CD_Prompt 3	1.78	0.63	1.36

4.3.2 Personalized Steering Wheel Design Solutions. First, the text requirement is entered into the configuration document to retrieve the most appropriate design solution, as shown in Fig. 9. We used the Word2Vec tool to calculate the cosine similarity between the user requirement and 10 configuration elements. Configuration element ID 7 has the highest similarity to the user's preferences, making it the most suitable option. The retrieved configuration element for the user's requirement is shown in Fig. 10(*a*). After presenting the design solution, the system initiates a user satisfaction feedback interface, as shown in Fig. 7(*c*). If the user is not satisfied with the configured design option, the system uses the text-to-3D generative model to generate a new design result, as shown in Fig. 10(*b*).

Table 4 Average texture LPIPS score [36] for generated texture fidelity evaluation

Type	Ours	Meshy AI	Hunyuan3D
LPIPS_Prompt 1	0.67	0.57	0.63
LPIPS_Prompt 2	0.83	0.64	0.61
LPIPS_Prompt 3	0.79	0.86	0.91

5 User Study

We conducted an empirical investigation to assess the effectiveness of the proposed AIGC-enabled personalized design methodology in improving user experience.

5.1 Experiment Setup. We initially hypothesize that the proposed AIGC-enabled immersive design system (System A) will enhance personalization, increase user engagement, facilitate user interaction experiences, and subsequently improve user satisfaction compared to the conventional configuration method (System B).

In our study, the null hypothesis H_0 and the alternative hypothesis H_1 are distinguished specifically by the difference in mean outcome scores for the four user experience indicators between System A and System B. Our statistical inference is based on comparing these mean scores: H_0 states that there is no significant difference, or that System B performs better than or equal to System A in terms of the mean scores on the four indicators and H_1 posits that System A achieves higher mean scores than System B on these indicators.

The required sample size was determined using G*Power 3.1. We performed a prior power analysis for a one-tailed paired *t*-test, with an alpha error probability (α) of 0.05, a statistical power (1-β) of 0.9, and an assumed medium effect size (Cohen's d = 0.5). The analysis indicated that at least 20 participants were needed to detect a statistically significant difference between the two systems. Therefore, we recruited 20 participants aged 21–35, all graduate students from diverse fields. Each participant was engaged in a task to create a personalized steering wheel design using both System A and System B in a randomized order. In System A, participants used voice input to specify their design requirements within an immersive environment, with the option to use a pretrained text-to-3D generative model for further personalization. In contrast, System B limited participants to selecting from predefined configuration options without generating new

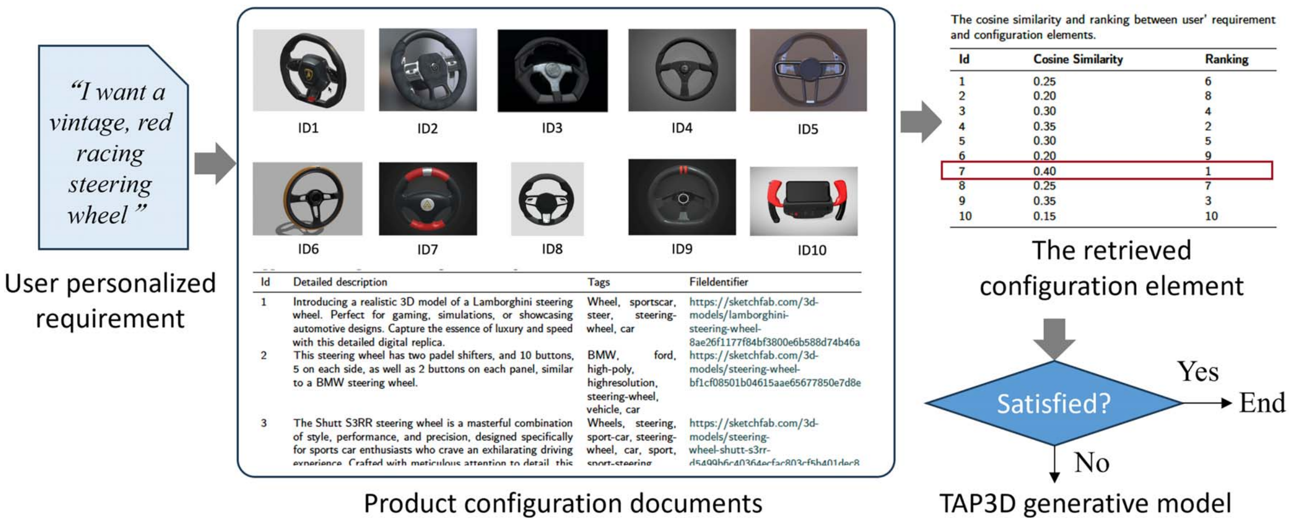

Fig. 9 The retrieval process of steering wheels

(*a*)

(*b*)

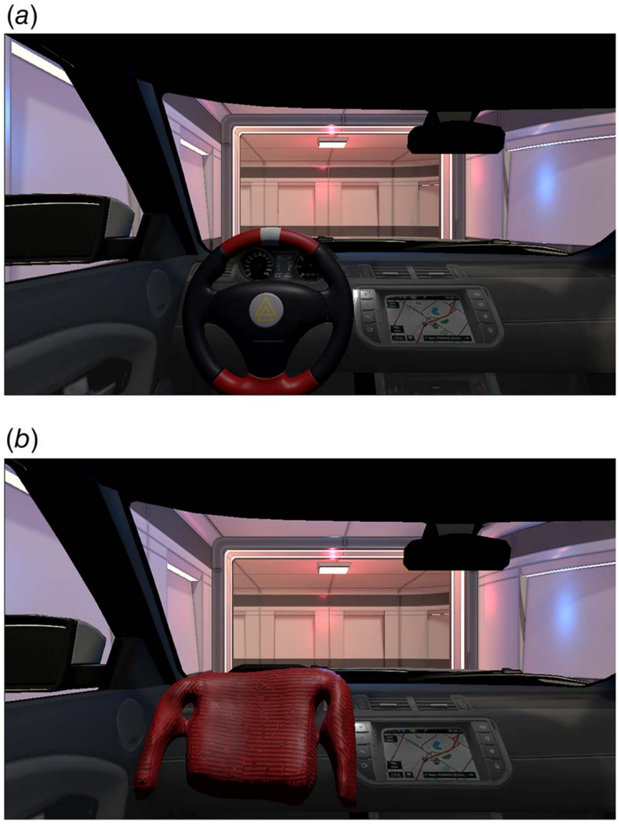

Fig. 10 (*a*): The retrieved personalized steering wheel design option from the configuration document. (*b*) The generated personalized steering wheel design solution from a text-to-3D generative model.

designs. After completing experimental tasks, participants filled out questionnaires to evaluate the two systems.

5.2 Metrics. To evaluate the impact of the proposed AIGC-enabled immersive design system, a 7-point Likert scale, ranging from 1 (not at all satisfactory) to 7 (very satisfactory), was employed to quantify the participants' responses. The evaluation criteria were divided into four key metrics: personalization degree, user engagement, interaction experience, and user satisfaction, as illustrated in Fig. 11. To provide a clearer understanding of these metrics, the specific explanations are outlined below:

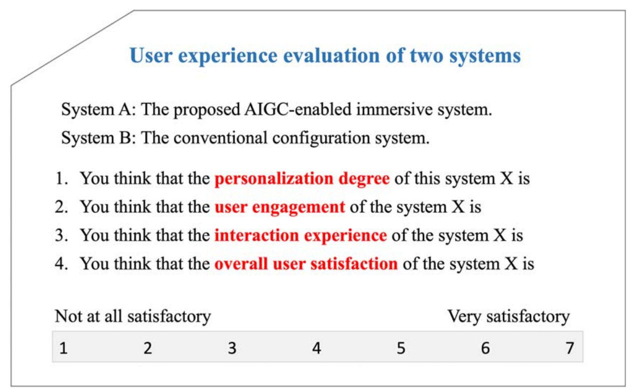

Fig. 11 The detailed questions design for user experience evaluation of the proposed system and the conventional configuration system

Q1: The degree of personalization refers to how well the method of providing personalized designs aligns with individual user needs.

Q2: User engagement assesses participants' sense of involvement and commitment throughout the personalized design process.

Q3: Interaction experience is evaluated based on the willingness, intuitiveness, and enjoyment during the interaction process.

Q4: User satisfaction measures the overall performance of the two systems.

5.3 Results. All participants accomplished the comparative experiment, successfully obtained a personalized steering wheel design within the VR environment. The mean values (M) of the four metrics for each system are illustrated in Fig. 12. System A outperformed System B across all evaluated metrics. The one-tail paired-sample t-test was used for the significance test, with a significance level set at 0.01. The detailed results are shown in Table 5.

Personalization degree: System A had a mean score M of 6.1 with a standard deviation (SD) of 0.738, while System B had M = 3.9 and SD = 0.994. The t-test resulted in t (19) = -4.975 with a significance level of $P < 0.001$. This indicates a rejection

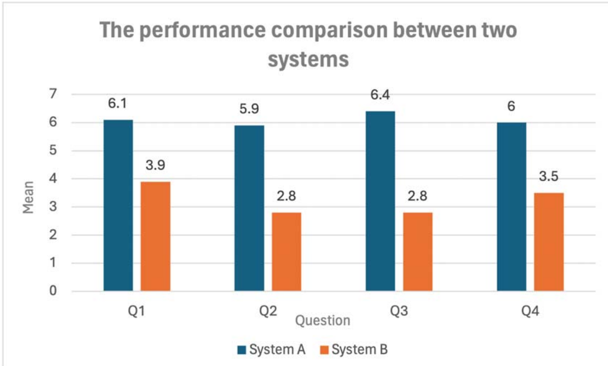

Fig. 12 The line chart shows the performance comparison between System A and System B

Table 5 The experiment analysis results of user study

	System A		System B			
Metrics	*M*	SD	*M*	SD	*t*	*P*
Q1	6.1	0.738	3.9	0.994	−4.975	$3.824 \times 10^{-04***}$
Q2	5.9	0.876	2.8	1.135	−6.432	$6.026 \times 10^{-05***}$
Q3	6.4	0.516	2.8	1.370	−9.0	$4.269 \times 10^{-06***}$
Q4	6.0	0.471	3.5	0.972	−7.319	$2.236 \times 10^{-05***}$

M = mean; SD = standard deviation.
$***P < 0.001$.

of the null hypothesis, suggesting that System A provides significantly more personalized solutions than System B.

User engagement: System A scored $M = 5.9$, SD = 0.876; System B scored $M = 2.8$, SD = 1.135. This analysis produced a *t*-value of $t(19) = -6.433$ and $P < 0.001$. This result leads to a rejection of the null hypothesis, indicating that participants demonstrated significantly greater engagement in the personalized design process when using the proposed System A compared to the traditional System B.

Interaction experience: System A had $M = 6.4$, SD = 0.516, while System B had $M = 2.8$, SD = 1.370. The result $t(19) = -9$, $P < 0.001$, shows a significantly better interaction experience with System A. This supports the idea that immersive environments improve user interaction, consistent with previous studies [37].

Overall user satisfaction: System A's $M = 6.0$, SD = 0.471 compared to System B's $M = 3.5$, SD = 0.972. The *t*-test result $t(19) = -7.319$, $P < 0.001$, indicates significantly higher satisfaction with System A.

6 Discussion

Despite demonstrating the effectiveness of the proposed system, several limitations remain. First, a notable limitation is the limited availability of high-quality 3D datasets suitable for training text-to-3D generative models. This scarcity of data directly impacts the fidelity and generalization capability of the generated output. Moreover, the current implementation relies solely on publicly available datasets, rather than proprietary datasets from specific manufacturers. Consequently, the generated designs may lack product-specific details and fail to align seamlessly with existing product families, highlighting the need for more comprehensive, diverse, and domain-specific 3D datasets to enhance the model's practical applicability and performance in industrial contexts. Second, in our evaluation section, System B was selected as the baseline due to its widespread use as a conventional configuration system in industry and research. However, System B is limited in terms of flexibility and personalization, as it typically offers only a fixed set of options and lacks the ability to generate novel designs or adapt to unique user requirements. In contrast, our proposed methodology leverages AIGC-enabled generation and immersive interaction, allowing users to actively participate in the design process, receive timely feedback, and explore a broader design space. This approach addresses the limitations of traditional systems by enabling greater personalization and creative freedom. Finally, the current system requires approximately two minutes per 3D shape generation. While this fails short of real-time performance standards typical in VR or gaming environments, it is considered acceptable within the context of early-stage design ideation, where fast but not instantaneous feedback is sufficient to support creative decision-making.

To address the limitations identified in our current work, future research should prioritize the creation or organization of multiple and high-quality 3D datasets, potentially through collaborations with industry partners to access proprietary data. Incorporating these proprietary datasets would enable generative models to learn product-specific features, thereby enhancing compatibility with existing product families. Furthermore, implementing advanced data augmentation techniques and exploring cross-domain adaptation methods, such as transferring knowledge from 2D images to 3D models, could mitigate the situation of data scarcity. Additionally, once the user has generated the model through text input to obtain a preliminary design, it becomes particularly important to enable interactive modifications to finalize the detailed design. By pursuing these directions, future research can overcome current limitations and significantly advance the field of text-to-3D generative models.

7 Conclusion

As personalization becomes increasingly important in product design, innovation strategies are essential to meet diverse user requirements. This study introduces an AIGC-enabled personalized design methodology that integrates a retrieval model and a text-to-3D generative model within an immersive environment. A case study on personalized steering wheel design was used to demonstrate the feasibility of this methodology. A user study comparing our proposed system with a conventional configuration system indicates significant improvements in personalization degree, user engagement, interaction experience, and user satisfaction.

Future work will address current system development limitations, particularly the performance of the text-to-3D generative model. Future research will explore the potential of using explainable AI and knowledge infusion techniques to enhance personalized design through natural language interactions. These approaches could improve user understanding and trust in the design process, leading to more tailored and contextually relevant design outcomes.

Acknowledgment

This work was mainly supported by funding from the National Natural Science Foundation of China (No. 52422514), the State Key Laboratory of Intelligent Manufacturing Equipment and Technology (Huazhong University of Science and Technology, IMETKF2024010), and the Shenzhen Science and Technology Innovation Commission (JCYJ20230807140407016).

Conflict of Interest

There are no conflicts of interest.

Data Availability Statement

The datasets generated and supporting the findings of this article are obtainable from the corresponding author upon reasonable request.

References

[1] Wang, Z., Zhang, M., Sun, H., and Zhu, G., 2016, "Effects of Standardization and Innovation on Mass Customization: An Empirical Investigation," Technovation, **48**, pp. 79–86.

[2] Wang, T., Fan, J., and Zheng, P., 2024, "An LLM-Based Vision and Language Cobot Navigation Approach for Human-Centric Smart Manufacturing," J. Manuf. Syst., **75**, pp. 299–305.

[3] Maddikunta, P. K. R., Pham, Q. -V., Prabadevi, B., Deepa, N., Dev, K., Gadekallu, T. R., Ruby, R., and Liyanage, M., 2022, "Industry 5.0: A Survey on Enabling Technologies and Potential Applications," J. Ind. Inf. Integr., **26**, p. 100257.

[4] Minnoye, A. L., Tajdari, F., Doubrovski, E. L., Wu, J., Kwa, F., Elkhuizen, W. S., Huysmans, T., and Song, Y., 2022, "Personalized Product Design Through Digital Fabrication," International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, St. Louis, MO, Aug. 14–17.

[5] Guo, F., Qu, Q.-X., Nagamachi, M., and Duffy, V. G., 2020, "A Proposal of the Event-Related Potential Method to Effectively Identify Kansei Words for Assessing Product Design Features in Kansei Engineering Research," Int. J. Ind. Ergon., **76**, p. 102940.

[6] Brown, D. C., 1998, "Defining Configuring," Artif. Intell. Eng. Des. Anal. Manuf., **12**(4), pp. 301–305.

[7] Wang, T., Zheng, P., Li, S., and Wang, L., 2024, "Multimodal Human–Robot Interaction for Human-Centric Smart Manufacturing: A Survey," Adv. Intell. Syst., **6**(3), p. 2300359.

[8] Chen, L., Jing, Q., Tsang, Y., Wang, Q., Sun, L., and Luo, J., 2024, "Designfusion: Integrating Generative Models for Conceptual Design Enrichment," ASME J. Mech. Des., **146**(11), p. 111703.

[9] Jiang, C., Huang, R., and Shen, T., 2025, "Generative Artificial Intelligence-Enabled Conceptualization: Charting Chat-Generative Pre-Training Transformer's Impacts on Sustainable Service Design Thinking With Network-Based Cognitive Maps," ASME J. Comput. Inf. Sci. Eng., **25**(2), p. 021006.

[10] Li, X., Wang, Y., and Sha, Z., 2022, "Deep Learning of Cross-Modal Tasks for Conceptual Design of Engineered Products: A Review," International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, St. Louis, MO, Aug. 14–17.

[11] Ren, M., and Zheng, P., 2024, "Towards Smart Product-Service Systems 2.0: A Retrospect and Prospect," Adv. Eng. Inform., **61**, p. 102466.

[12] Dong, W., Li, S., and Zheng, P., 2025, "Toward Embodied Intelligence-Enabled Human–Robot Symbiotic Manufacturing: A Large Language Model-Based Perspective," ASME J. Comput. Inf. Sci. Eng., **25**(5), p. 050801.

[13] Suh, A., and Prophet, J., 2018, "The State of Immersive Technology Research: A Literature Analysis," Comput. Hum. Behav., **86**, pp. 77–90.

[14] Huang, H.-M., Rauch, U., and Liaw, S.-S., 2010, "Investigating Learners' Attitudes Toward Virtual Reality Learning Environments: Based on a Constructivist Approach," Comput. Educ., **55**(3), pp. 1171–1182.

[15] Hashim, M. E. A., Mustafa, W. A. W., Prameswari, N. S., Ghani, M. M., and Hanafi, H. F., 2024, "Revolutionizing Virtual Reality With Generative AI: An In-Depth Review," J. Adv. Res. Comput. Appl., **30**(1), pp. 19–30.

[16] Ratican, J., Hutson, J., and Wright, A., 2023, "A Proposed Meta-Reality Immersive Development Pipeline: Generative AI Models and Extended Reality (XR) Content for the Metaverse," Journal of Intelligent Learning Systems and Applications, **15**(01), pp. 24–35.

[17] Zhang, L., Pan, J., Gettig, J., Oney, S., and Guo, A., 2024, "VRCopilot: Authoring 3D Layouts With Generative AI Models in VR," Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology, Pittsburgh, PA, Oct. 13–16, pp. 1–13.

[18] Rasch, J., Töws, J., Hirzle, T., Müller, F., and Schmitz, M., 2025, "Creepycocreator? Investigating AI Representation Modes for 3D Object Co-Creation in Virtual Reality," Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems., Yokohama, Japan, Apr. 26–May 1, pp. 1–14

[19] Stemasov, E., Wagner, T., Gugenheimer, J., and Rukzio, E., 2020, "Mix&match: Towards Omitting Modelling Through in-Situ Remixing of Model Repository Artifacts in Mixed Reality," Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, Apr. 25–30 , pp. 1–12.

[20] Weng, S. C.-C., Chiou, Y. -M., and Do, E. Y.-L., 2024, "Dream Mesh: A Speech-to-3D Model Generative Pipeline in Mixed Reality," 2024 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR), Los Angeles, CA, Jan. 17–19, IEEE, pp. 345–349.

[21] Behravan, M., Matković, K., and Gračanin, D., 2025, "Generative AI for Context-Aware 3D Object Creation Using Vision-Language Models in Augmented Reality," 2025 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR), Lisbon, Portugal, Jan. 27–29, IEEE, pp. 73–81.

[22] Behravan, M., and Gračanin, D., 2025, "From Voices to Worlds: Developing an AI-Powered Framework for 3D Object Generation in Augmented Reality," 2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Saint Malo, France, Mar. 8–12, IEEE, pp. 150–155.

[23] Xing, Y., Liu, Q., Wang, J., and Gomez-Zara, D., 2024, "sMoRe: Enhancing Object Manipulation and Organization in Mixed Reality Spaces With LLMS and Generative AI," arXiv preprint arXiv:2411.11752.

[24] Truong, V. T., and Le, L. B., 2024, "Text-Guided Real-World-to-3D Generative Models With Real-Time Rendering on Mobile Devices," 2024 IEEE Wireless Communications and Networking Conference (WCNC), Dubai, United Arab Emirates, Apr. 21–24, IEEE, pp. 1–6.

[25] Liu, Z., Dai, P., Li, R., Qi, X., and Fu, C.-W., 2023, "Dreamstone: Image as a Stepping Stone for Text-Guided 3D Shape Generation," IEEE Trans. Pattern. Anal. Mach. Intell., **45**(12), pp. 14385–14403.

[26] Li, X., Sun, Y., and Sha, Z., 2025, "LLM4CAD: Multimodal Large Language Models for Three-Dimensional Computer-Aided Design Generation," ASME J. Comput. Inf. Sci. Eng., **25**(2), p. 021005.

[27] Rong, X., 2016, "word2vec Parameter Learning Explained," arxiv. .

[28] Wei, J., Wang, H., Feng, J., Lin, G., and Yap, K.-H., 2023, "TAPS3D: Text-Guided 3D Textured Shape Generation from Pseudo Supervision," 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, June 17–24.

[29] Gao, J., Shen, T., Wang, Z., Chen, W., Yin, K., Li, D., Litany, O., Gojcic, Z., and Fidler, S., 2022, "GET3D: A Generative Model of High Quality 3D Textured Shapes Learned From Images," Advances in Neural Information Processing Systems, New Orleans, LA, Nov. 28–Dec. 9.

[30] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., and Sastry, G., 2021, "Learning Transferable Visual Models From Natural Language Supervision," International Conference on Machine Learning, Virtual Event, July 18–24, PMLR,pp. 8748–8763.

[31] Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z. et al., 2015, "ShapeNet: An Information-Rich 3D Model Repository," arXiv preprint arXiv:1512.03012.

[32] Deitke, M., Schwenk, D., Salvador, J., Weihs, L., Michel, O., VanderBilt, E., Schmidt, L., Ehsani, K., Kembhavi, A., and Farhadi, A., 2023, "Objaverse: A Universe of Annotated 3D Objects," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13142–13153.

[33] Yang, X., Shi, H., Zhang, B., Yang, F., Wang, J., Zhao, H., Liu, X. et al., 2024, "Hunyuan3D 1.0: A Unified Framework for Text-to-3D and Image-to-3D Generation," arXiv preprint arXiv:2411.02293.

[34] Hessel, J., Holtzman, A., Forbes, M., Bras, R. L., and Choi, Y., 2021, "Clipscore: A Reference-Free Evaluation Metric for Image Captioning," arXiv preprint arXiv:2104.08718.

[35] Borgefors, G., 1988, "Hierarchical Chamfer Matching: A Parametric Edge Matching Algorithm," IEEE Trans. Pattern Anal. Mach. Intell., **10**(6), pp. 849–865.

[36] Ding, K., Ma, K., Wang, S., and Simoncelli, E. P., 2020, "Image Quality Assessment: Unifying Structure and Texture Similarity," IEEE Trans. Pattern Anal. Mach. Intell., **44**(5), pp. 2567–2581.

[37] Yin, Y., Zheng, P., Li, C., Cong, J., and Pang, Y. M., 2023, "An Empirical Study of an MR-Enhanced Kinematic Prototyping Approach for Articulated Products," Adv. Eng. Inform., **58**, p. 102203.