From Automation to Intelligence

An AI agent for spatial reasoning and 3D co-design

Ximing Zhong[1], Yingkai Li[2], Pia Fricker[3], Xiange Zhang[4]
[1,3]Aalto University [2]Politecnico di Milano [4]University of Washington
[1]{ximing.zhong|pia.fricker}@aalto.fi [2]yingkai.li@mail.polimi.it [3]xiange@uw.edu

Spatial understanding and reasoning are core cognitive abilities in architectural design, yet existing AI tools remain limited to automation without interpretability or goal-driven logic. This study introduces a 3D Co-Design Assistant that integrates spatial perception, reasoning, and modeling via a Large Language Model (LLM)-driven agent framework. Operating through a real-time perception–reasoning–action loop, the system enables AI to generate design suggestions and autonomously execute modeling commands in a 3D environment. Unlike previous black-box generative models, the assistant produces interpretable design logic grounded in structured scene data and user input. Its capabilities are evaluated through controlled human judgment and co-creation experiments, where participants often misidentified AI suggestions as human-generated, and rated many proposals as rational and contextually appropriate—demonstrating early signs of human-like spatial cognition. By bridging semantic reasoning with executable modeling behaviors, the proposed framework offers a novel approach to AI-assisted design beyond conventional automation, highlighting the potential of intelligent agents to support architects in iterative spatial thinking, stimulate creative alternatives, and facilitate the integration of AI into early-stage architectural workflows.

Keywords: *AI-Driven Design, 3D Co-Design, Spatial Reasoning, Large Language Models, Agent-Based Design.*

INTRODUCTION

Spatial understanding and spatial reasoning are essential cognitive abilities that architects must develop throughout their training, underpinning the entire architectural design process. Spatial understanding involves analyzing spatial information and user needs, while reasoning concerns the formulation and inference of design objectives. Traditionally, architects collect user needs through interviews, surveys, or observations and abstract them into spatial understanding, followed by iterative reasoning to refine 3D models until goals are met an inherently intelligent process.

In recent years, advances in AI and digital tools have sparked efforts to endow machines with similar understanding and reasoning abilities–not only to accelerate modeling, but to transform AI into an intelligent assistant for creative spatial design. However, this remains a big challenge. Current workflows rely on textual and graphical representations to convey design intent, while the underlying reasoning still depends on human judgment. This cross-modal cognitive chain is complex and implicit, making it difficult to digitize and encode into generative systems.

Tools such as CAD, BIM, and environmental analysis software (e.g., lighting, thermal, airflow) have improved efficiency and precision. Generative design methods further automate certain tasks using rule-based generation. However, these tools still require human interpretation and reasoning, lacking the ability to autonomously perceive or make spatial decisions. While these tools offer automated performance and parametric flexibility, they lack the ability to engage in active spatial interpretation or decision-making – limiting their function to automation rather than intelligent collaboration.

Recently, data-driven 3D generation approaches such as GANs, 3DGNNs, and 3D-GANs have emerged in AI research (Zhong et al., 2022). These methods generate 3D layouts by learning statistical correlations from existing datasets but lack explicit spatial reasoning or interpretable design motivations. Consequently, architects find it difficult to trust or reliably adopt their outputs. These models operate primarily as statistical approximators, automating form generation without engaging in goal-directed spatial reasoning – revealing the limitations of current AI design tools as merely automated rather than intelligent.

In parallel, emerging cross-modal generative AI technologies such as Text-to-3D models have begun to bridge natural language input and 3D model generation, exhibiting preliminary semantic-to-spatial alignment capabilities (Xu et al., 2023). Yet, these methods primarily map textual descriptions to 3D forms in an end-to-end manner, lacking the ability to model spatial structure, layout logic, and explicit reasoning chains. As a result, they fall short of achieving human-like spatial understanding and reasoning in the design process (Liu et al., 2024; Sun et al., 2023). The integration of a complete understanding-reasoning-modeling process into design tools remains an open challenge.

Bridging this gap requires a fundamental shift–from viewing AI as an automation tool to positioning it as an intelligent agent capable of reasoning, interpreting, and collaborating in design tasks. Against this backdrop, the concept of the AI agent has revealed great potential for enabling more intelligent design tools. An AI agent refers to an artificial intelligence system capable of perceiving its environment, understanding the contextual information, reasoning about problem-solving pathways, and autonomously planning, acting, and utilizing external tools. It possesses a closed-loop ability of continuous perception-reasoning-action in dynamic environments, enabling human-like decision-making and collaboration (Russell & Norvig, 2020). This capability precisely addresses the needs of architects in the processes of environmental understanding and spatial reasoning.

Building on this, we propose an innovative interactive AI-driven spatial design framework–the 3D Co-Design Assistant. The primary objective of this study is to enable AI systems to acquire spatial understanding and reasoning abilities, thereby assisting designers in completing 3D spatial design tasks more intelligently.

The key contributions of this paper are summarized as follows: (1) We present an AI design assistant capable of human-like spatial understanding and reasoning for 3D modeling, which can autonomously analyze environmental context and generate logically coherent and interpretable design proposals. (2) We develop a modeling command and operation library tailored for AI control, enabling Large Language Models to translate semantic reasoning into executable 3D modeling actions within design environments. (3) We propose and implement a full-fledged Co-Design framework that integrates perception, reasoning, and modeling, advancing AI from automating modeling tasks to intelligently assisting human design decisions and reasoning processes.

RELETED WORK

AI data-driven 3D model generation methods

In recent years, AI data-driven 3D generation methods have gained significant attention in the field of architectural design. Representative approaches include BuildingGAN, BuildingGNN, and 3D Diffusion models (Chang et al., 2021; Sebestyen et al., 2023). These methods typically rely on large-scale 3D model datasets to learn the statistical distribution of spatial morphology, enabling the automatic generation of 3D spatial solutions under given conditions such as site constraints, boundaries, and functional requirements. For instance, upon receiving site constraint inputs, these models can generate 3D layouts or forms that align with the distribution of the training data.

However, these methods primarily depend on data distribution fitting, lacking reasoning chains and interpretability in their generation mechanisms. As a result, they fail to comprehend the design intent underlying the input constraints and are incapable of conducting causal reasoning or dynamic adjustments in the spatial decision-making process. Consequently, they still exhibit certain limitations regarding the rationality and credibility of their generated results.

Text-to-3D cross-modal generation methods

Text-to-3D technology aims to achieve automatic 3D morphology generation from semantic descriptions by establishing a cross-modal mapping between natural language and 3D models. Representative approaches include DreamFusion, Luma AI, and Shap-E (Poole et al., 2022), which typically leverage diffusion models or neural rendering techniques to transform user-provided text prompts into 3D models that align with the given descriptions (Huang et al., 2024). These methods demonstrate AI's initial ability to associate natural language with spatial generation.

However, these approaches exhibit certain limitations. On the one hand, their generation process primarily follows an end-to-end semantic-to-morphology mapping, lacking considerations for spatial functionality, structural logic, and layout reasoning—key dimensions of architectural design. On the other hand, current methods struggle to enable real-time interaction and iterative optimization in human-AI collaborative design, as they lack multi-step reasoning and decision-making chains tailored to design intent. This constraint significantly limits their practical applicability in architectural design scenarios.

Agent-based architectures for 3D spatial reasoning

With the advancement of Large Language Models (LLMs) and cross-modal reasoning technologies, the AI Agent architecture is emerging as a key pathway toward intelligent spatial design. An AI Agent is a system equipped with environmental perception, task understanding, reasoning capabilities, autonomous planning, and action execution, enabling it to perform multi-step reasoning and interactive tasks within dynamic environments (Tan et al., 2024; Zhou et al., 2024). Its core objective is to simulate the "perception–thinking–action" closed-loop process that human designers employ, facilitating the complete workflow from task comprehension to 3D solution generation. In the field of architectural spatial generation, a representative 3D generative agent system is the BuildingAgent architecture proposed by Zhong et al. (Zhong, Koh & Fricker, 2023). This framework combines LLMs' semantic understanding (e.g., GPT-4) with GNN-based layout reasoning, enabling AI to process 3D geometric contexts and natural language design tasks. However, it still lacks essential elements of human-like spatial understanding and reasoning. Its decision-making relies on pretrained layout patterns without goal-driven semantic interpretation, leaving the underlying logic opaque and the system incapable of autonomously understanding design objectives or reasoning through spatial decisions.

To overcome these limitations, this paper proposes a novel Agent architecture that integrates 3D generation and spatial reasoning capabilities, aiming to address existing limitations in spatial cognitive intelligence, reasoning chain construction, and design interpretability, thereby advancing the intelligent evolution of AI in architectural spatial design.

METHODOLOGY

This study proposes a modular AI-assisted construction framework consisting of three core modules – Perception, Thinking, and Action – facilitating real-time human-AI collaboration in a 3D environment (Figure 1). This system provides continuous design suggestions and operational feedback.

The Perception Module acquires spatial data from the current 3D scene, capturing surrounding structures and architectural attributes. Conceptually, its goal is to convert the spatial environment into a machine-readable representation. Technically, this module retrieves real-time scene data from the Unreal Engine 5 (UE5) environment, including object names, dimensions, positions, and semantic categories. It parses both visual geometry and semantic tags through UE ' s reflection and data interface system and outputs a structured 3D representation (3D Structure) in a JSON-like format. This serves as the foundation for subsequent reasoning and modeling operations. The Perception Module includes two types of data parsing processes: one for geometric features (mesh bounds, coordinates, object types) and another for semantic attributes (style, category, spatial relationships), as illustrated in Figure 2.

Based on this structured input, the Thinking Module performs scene analysis and reasoning. The objective of this module is to interpret the spatial context and generate intelligent, high-level design suggestions. It utilizes Large Language Models (LLMs), specifically GPT-4o, and follows three reasoning principles: (1) Rigorous Thinking - every suggestion must be accompanied by an explanation; (2) Contextual Relevance - suggestions must be directly informed by the current architectural and environmental context; (3) Structured Output - clear separation between reasoning and design proposal. No fine-tuning was conducted; all reasoning was achieved through prompt engineering using structured templates (Figure 2).

Once a suggestion is accepted by the user, the Action Module, which also powered by LLMs translates it into executable modeling operations.

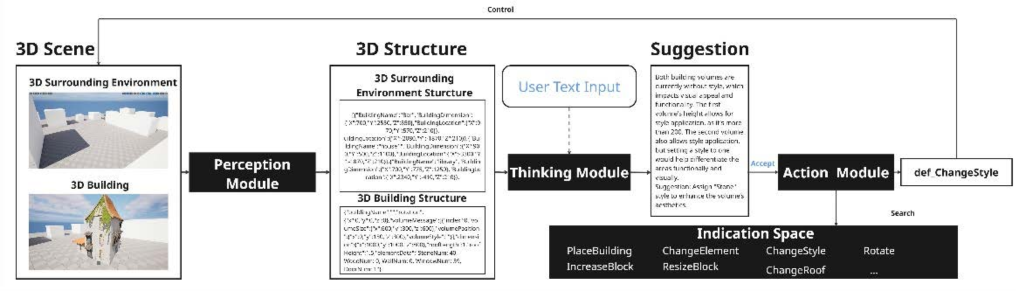


Figure 1
System framework of the 3D co-design

Figure 2
Perception and thinking module workflow of the 3D co-design assistant

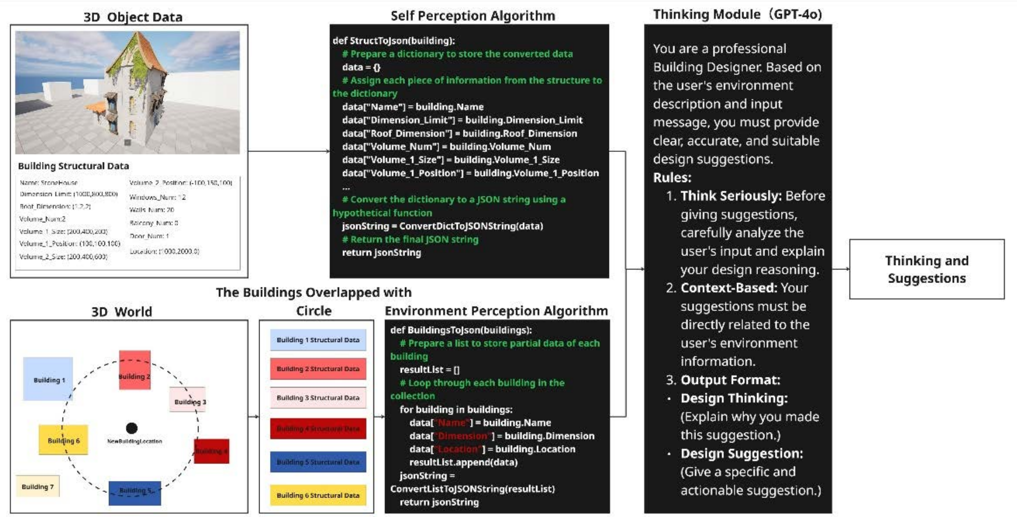
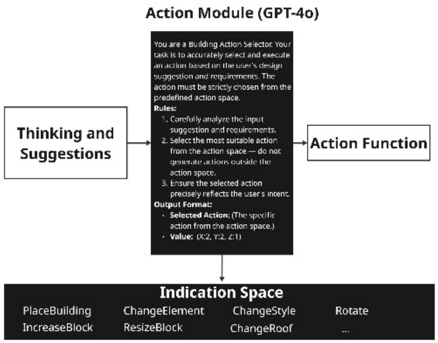


Figure 3
Action module : AI selects commands to model behaviors based on reasoning results



This module functions as an architectural action selector, converting semantic design intent into concrete commands within a predefined action space. And all actions is rule based. The action categories include PlaceBuilding, IncreaseBlock, ResizeBlock, Rotate, ChangeStyle, ChangeRoof, and ChangeElement. Actions are selected through rule-based matching of suggestion content to available commands. To maintain consistency and control, the system is constrained to operate strictly within this predefined library. Each selected action is output as a structured operation with spatial parameters (X, Y, Z), which is then executed via callable Blueprint functions in UE5 (e.g., def_ChangeStyle). Figure 3 illustrates this structured selection and execution process.

EXPERIMENTS

Human judgment experiment

To examine users' perception and evaluation of AI-generated suggestions in collaborative architectural design, a controlled human judgment experiment was conducted. The study addressed two questions: whether users could accurately identify the source of design suggestions—AI or human—and how they evaluated the rationality and practicality of these suggestions.

A total of 20 participants were recruited, including senior architecture students and junior practitioners, all with prior exposure to conceptual design tasks and moderate familiarity with AI-based tools. Each participant completed the experiment independently using a standardized web interface.

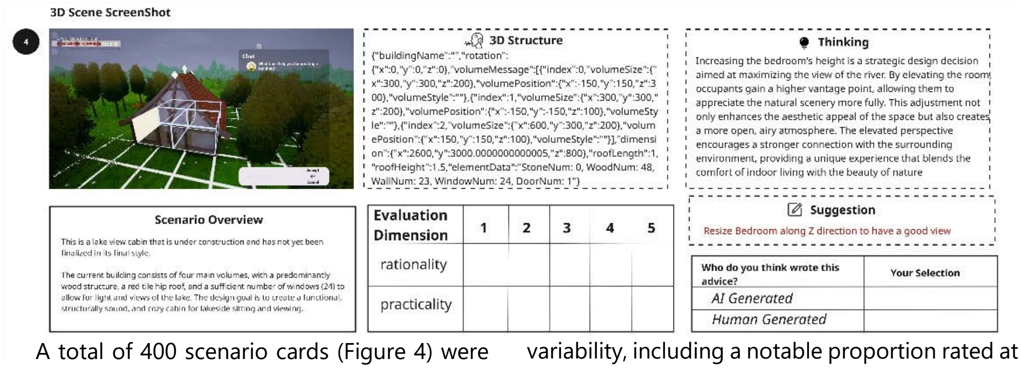


Figure 4
Example of a scene card used in the source identification experiment

A total of 400 scenario cards (Figure 4) were created. The suggestions were randomly sourced from either an AI system or a human designer, with the source concealed from participants. To minimize linguistic cues that could reveal the source, human suggestions were standardized to match the neutral and consistent style of AI outputs. Each card also included basic information about the building structure and design scenario to support user evaluation.

Participants were asked to complete two tasks per card: (1) Identify whether the suggestion was generated by AI or a human. (2) Rate its rationality and practicality on a five-point Likert scale. A total of 1,000 valid responses were collected.

The results showed that participants achieved an overall identification accuracy of 63.5%, exceeding the random baseline of 50% (Figure 5(a)). Accuracy for AI-generated suggestions reached 71.4%, while that for human-generated suggestions was 59.2%. As shown in the confusion matrix (Figure 5(b)), human suggestions were more frequently misidentified as AI-generated (40.8%) than vice versa (28.6%), indicating cognitive ambiguity influenced by the human-like expression of AI outputs.

In subjective evaluations, human-generated suggestions consistently received higher and more stable scores across both dimensions. Their rating distribution (Figure 6) was more concentrated, with few extreme low scores. In contrast, AI suggestions showed greater variability, including a notable proportion rated at the lowest score (1 point), reflecting current limitations in spatial reasoning and contextual awareness.

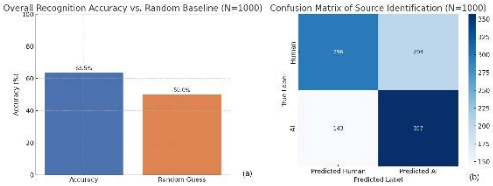


Figure 5
Recognition accuracy(a) and confusion matrix(b) of source identification (N = 1,000)

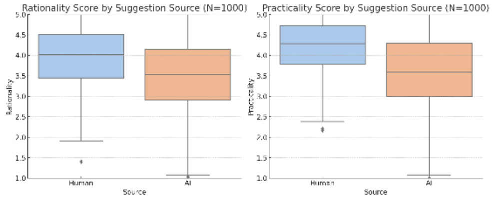


Figure 6
Subjective evaluation scores of suggestion quality (N=1000)

Nonetheless, the results also reveal the potential of AI in collaborative design. Several AI-generated suggestions were rated as highly as those by human designers, demonstrating the system's capacity to contribute valuable and novel inputs under specific conditions.

Overall, these findings suggest that while AI-generated suggestions can enhance design diversity and stimulate alternative thinking, their quality inconsistency and contextual limitations necessitate human oversight. In this context, AI is more suitable as an auxiliary design partner rather than an autonomous decision-maker.

3D Co-Creation Experiment

To assess the performance of the AI assistant in collaborative architectural design tasks, we employed four evaluation indicators: task completion rate, suggestion acceptance rate, action response rate, and logical score. The task completion rate measured whether the system fulfilled predefined design objectives in each round. The suggestion acceptance rate indicated the proportion of AI-generated suggestions adopted by users, while the action response rate assessed whether accepted suggestions correctly triggered the intended modeling operations.

To evaluate the coherence, contextual relevance, and feasibility of each suggestion, we used a logical score based on a simplified three-point scale (1 = weak or irrelevant, 2 = partially logical, 3 = fully logical). This adjustment—compared to the five-point Likert scale used in the Human Judgement Experiment—was made to reduce cognitive load and facilitate faster, more consistent real-time assessment during interactive co-design sessions.

The 3D Co-Creation Experiment involved eight participants, including six architecture students and two early-career professionals. All participants had prior experience with 3D modeling tools and basic familiarity with AI-assisted design workflows. Each participant engaged in a structured session with the AI assistant, completing a series of collaborative design tasks. Suggestions were evaluated both during and after interaction, with emphasis placed on functional logic and contextual appropriateness. Task types spanned different spatial scales and complexities—namely, a community center, a lakeside residence, and a nighttime kiosk. To ensure consistency across sessions, all scenes and AI-generated suggestions were standardized.

In the community service center task (Figure 7), the AI assistant demonstrated clear understanding of design objectives and provided causally logical suggestions, such as volumetric division and south-facing window placement. The task completion rate, suggestion acceptance rate, and action response rate all reached 100%, with an average logical score of 3.0.

In the lakeside residence task (Figure 8(a)), the AI assistant exhibited contextual sensitivity to goals such as openness, daylighting, and view orientation. It proposed strategic suggestions including volume height adjustment and massing reorientation. The task recorded a suggestion acceptance rate and action response rate of 87.5%, with a logical score of 2.85.

In the nighttime kiosk task (Figure 8(b)), while the AI assistant completed all required modeling operations, performance fluctuations emerged in multi-round interactions. Issues such as repetitive suggestions and stylistic misjudgments led to a lower suggestion acceptance rate and action response rate of 62.5%, and a logical score of 2.6.

Across all three tasks, the system achieved an average task completion rate of 100%, a suggestion acceptance rate and action response rate of 83.3%, and an average logical score of 2.82. Beyond these quantitative outcomes, the experiment highlighted the AI assistant's emerging cognitive behaviors. The system demonstrated an ability to integrate spatial context and user intentions, forming reasoning chains and generating explainable design suggestions. This closed-loop process—combining perception, reasoning, and action—marks a shift from passive automation to proactive and interpretable design collaboration.

However, two key limitations were identified. First, the system exhibited contextual forgetting and goal drift during extended interactions, leading to a decline in logical consistency over time. Second, inconsistencies were observed between semantic intent and modeling

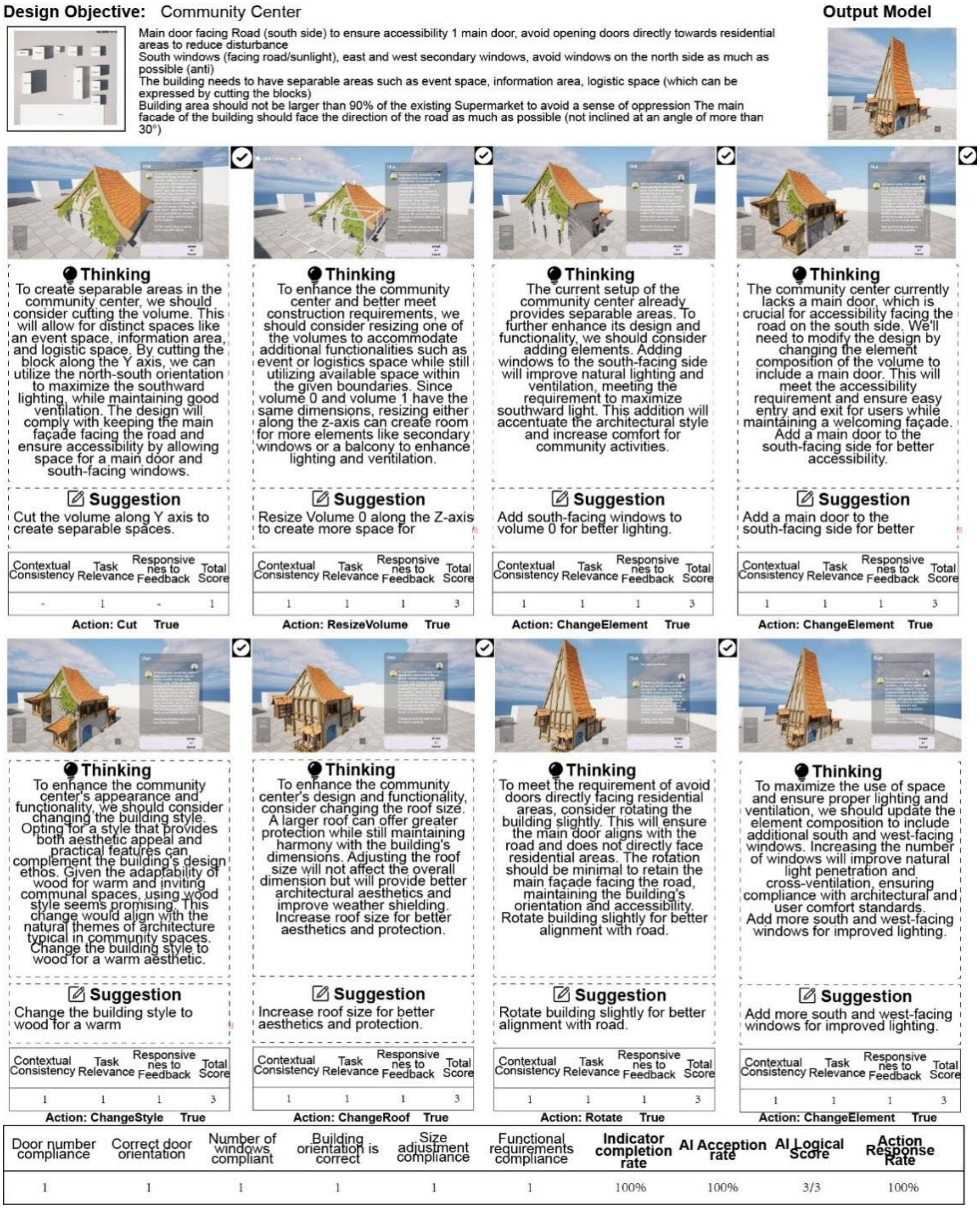
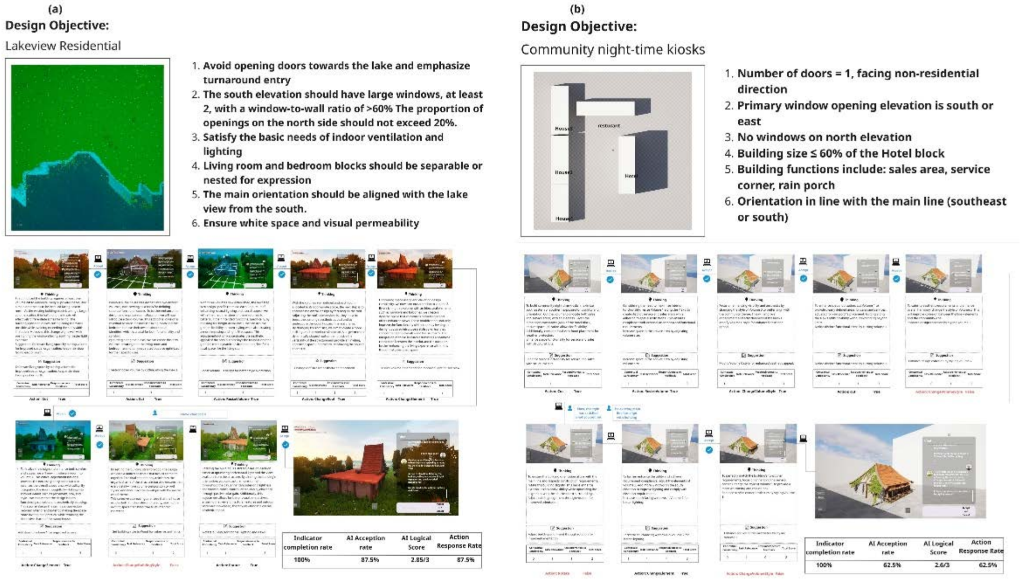


Figure 7
Co-design process and performance metrics for the community center task

Figure 8
Co-design process and performance metrics for Lake House(a) and Community Night-time Kiosks task(b)



execution, where clear suggestions occasionally failed to trigger appropriate actions. Addressing these issues will require improvements in long-term context retention and more robust semantic-action mapping to further support human-AI collaborative design.

LIMITATIONS AND FUTURE WORK

While the proposed 3D Co-Design Assistant demonstrates promising capabilities in spatial reasoning and collaborative design, the experiments revealed three major limitations.

First, the system exhibited contextual forgetting and goal drift during multi-round interactions. As design sessions progressed – particularly beyond the fourth round – its suggestions often became repetitive, diverged from initial objectives, or contradicted prior decisions. For example, in the Lakeside task, the assistant repeatedly recommended a "wood style" that had already been applied, indicating a lack of persistent memory and limited awareness of evolving design states.

Second, the system showed semantic-action inconsistency. In several cases, well-structured suggestions failed to trigger the appropriate modeling operations. For instance, the suggestion to "adjust the proportion of openings on the south facade" led to irrelevant or null actions, revealing weak alignment between high-level design intent and executable modeling behaviors.

Third, the system currently focuses on conceptual spatial modeling and lacks native support for BIM or AEC-specific data formats. Nonetheless, this limitation can be addressed through interoperability tools such as Rhino.Inside, Datasmith, and glTF export, which allow conversion of generated geometry into formats compatible with platforms like Revit and Rhino-Grasshopper.

To address these limitations, future development will focus on three directions: (1) integrating memory-augmented language models or structured memory graphs to improve long-term state tracking and dialogue coherence;

(2) enhancing semantic-action grounding through rule-based matching, supervised fine-tuning, or reinforcement learning; and (3) enabling deeper integration with BIM and parametric toolchains to support downstream workflows such as documentation, simulation, and construction.

More broadly, realizing AI as a genuine co-creator in architectural design requires systems that go beyond reactive assist.

REFERENCES

Xu, J., Wang, X., Cheng, W., Cao, Y.-P., Shan, Y., Qie, X. and Gao, S. (2023) 'Dream3d: Zero-shot text-to-3d synthesis using 3d shape prior and text-to-image diffusion models' in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20908-20918.

Zhong, X., Fricker, P., Yu, F., Tan, C. and Pan, Y. A (2022) 'Discussion on an Urban Layout Workflow Utilizing Generative Adversarial Network (GAN)-With a focus on automatized labeling and dataset acquisition' in *Proceedings of the International Conference on Education and Research in Computer Aided Architectural Design in Europe, 2022*. *eCAADe*, 583-592.

Sun, C., Han, J., Deng, W., Wang, X., Qin, Z. and Gould, S. (2023) '3d-gpt: Procedural 3d modeling with large language models', *arXiv preprint* arXiv:2310.12945.

Liu, D., Huang, X., Hou, Y., Wang, Z., Yin, Z., Gong, Y., Gao, P. and Ouyang, W. (2024) 'Uni3d-llm: Unifying point cloud perception, generation and editing with large language models', *arXiv preprint* arXiv:2402.03327.

Zhong, X., Koh, I. and Fricker, P. (2023) 'Building-GNN: Exploring a co-design framework for generating controllable 3D building prototypes by graph and recurrent neural networks' *in Proceedings of International Conference on Education and Research in Computer Aided Architectural Design in Europe, 2023*. *eCAADe*, 431-440.

Sebestyen, A., Özdenizci, O., Legenstein, R. and Hirschberg, U. (2023) 'Generating conceptual architectural 3d geometries with denoising diffusion models' in *Digital Design Reconsidered-Proceedings of the 41st Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe 2023)*, 451-460.

Poole, B., Jain, A., Barron, J. T. and Mildenhall, B. (2022) 'Dreamfusion: Text-to-3d using 2d diffusion' *arXiv preprint* arXiv:2209.14988.

Chang, K.-H., Cheng, C.-Y., Luo, J., Murata, S., Nourbakhsh, M. and Tsuji, Y. (2021) 'Building-gan: Graph-conditioned architectural volumetric design generation' in *Proceedings of the IEEE/CVF international conference on computer vision*, 11956-11965.

Huang, T., Zeng, Y., Zhang, Z., Xu, W., Xu, H., Xu, S., Lau, R. W. and Zuo, W. (2024) 'Dreamcontrol: Control-based text-to-3d generation with 3d self-prior' in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2024. 5364-5373.

Tan, W., Ding, Z., Zhang, W., Li, B., Zhou, B., Yue, J., Xia, H., Jiang, J., Zheng, L. and Xu, X. (2024) 'Towards general computer control: A multimodal agent for red dead redemption ii as a case study' *ICLR 2024 Workshop on Large Language Model (LLM) Agents*, 2024.

Zhou, X., Ran, X., Xiong, Y., He, J., Lin, Z., Wang, Y., Sun, D. and Yang, M. H. (2024) 'Gala3d: Towards text-to-3d complex scene generation via layout-guided generative gaussian splatting' *arXiv preprint* arXiv:2402.07207.