Conversational Application of Agentic Multimodal AI in Collaborative Architectural Design Environment

An architectural-focus AI design partner for early-stage design exploration

Lok Hang Cheung[1], Likai Wang[2] and Dongxue Lei[3]
[1,2]Department of Architecture, Xi'an Jiaotong-Liverpool University
[3]Ningbo University
[1]lokhang.cheung19@student.xjtlu.edu.cn, ORCID: 0009-0001-2911-3733, [2]wang.likai@outlook.com, ORCID: 0000-0003-4054-649X
[3]leidongxue@nbu.edu.cn, ORCID: 0009-0006-1634-837X

Abstract. Nowadays, most AI-assisted architectural design processes are limited to one-to-one interaction and one-directional design processes and require a long learning curve for applying sophisticatedly designed AI applications. Despite agentic AI systems showing promising potential in enhancing human-AI collaboration in creative scenarios, they are underexplored in architectural design. This paper proposes a conversational design framework for early-stage architectural design exploration, with a developed architectural-focus agentic multimodal AI system deployed on WeChat. This common messaging application allows group messaging as a platform for collaborative design. A case study is presented by a mixed-use renovation development in Suzhou; tutors and three undergraduate architecture students joined a WeChat group with the AI bot as a design team. Each student developed a design proposal, and three AI collaboration modes were observed, including performance-based, evaluation-based and designerly-based. The proposed framework and agentic AI tool enable designers to dedicate themselves to the design process with minimum learning curves. Collaborative environment integration allows more sophisticated designers, such as tutors, to enhance design conversation by asking follow-up questions. Students displayed mutual-learning AI-collaborated design methods by observing others' conversations in group chat. With the observed challenges and opportunities, the implications of research, AI education, and real-world practice are discussed.

Keywords. Multimodal AI, Agentic AI, Architectural Design Process, Conversational Design Process, Collaborative Design Environment

1. Introduction

AI applications, especially multimodal AIs, which could process both verbal and visual inputs, have shown great potential in the early-stage architectural design process for

ARCHITECTURAL INFORMATICS, Proceedings of the 30th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2025, Volume 1, 173-182.

design exploration, ideation, and conception. However, several problems are observed that might diminish its utility in real-world scenarios. First, as Bolojan et al. (2024) stated, AI application in the architectural design process nowadays is mostly a linear “input-output” relationship, which makes the design process reductionist instead of a divergent, open-ended process. They proposed a conversational loop between AIs and human designers during the architectural process, whereas it requires a sophisticated understanding of deep learning to employ the system, making it difficult for most designers to use. Second, the outputs produced by generative AIs are often blamed for having little relevance to the site context. Zhong et al. (2024) proposed a design workflow that could leverage the context-awareness of performative optimisation together with the exploration capabilities of generative AIs. However, there are limited iterations between the performative optimisation and AI applications, making the design process one-directional.

There are common challenges found in human-AI collaborative scenarios. Scholars such as Abedin et al. (2022) and Ibarrola et al. (2024) discussed how introducing agentic AI systems can enhance the collaborative design process between humans and AI. Such systems allow the AI application to plan a proper workflow to respond to the users comprehensively. However, this technique is underexplored in architectural design scenarios. In response, this study aims to explore the opportunities of introducing agentic workflow in an AI-collaborated design process for early-stage exploration. An AI bot is developed with the aforementioned agentic workflow technique. It is implemented so that after it receives a designer’s input, it first analyses the designer’s input and then actively provides multi-perspective responses such as web-search information and reference images, instead of a one-off reply merely based on the designer’s question. As such, the AI bot is acting as a suggestive, conversational design partner instead of a predictive/automatic “fast draftsman”, as described by Nicholas Negroponte (1976). In this research, "conversation" refers to Donald Schön (2017) statement that design is a reflective process between the designer and the design enquiry, as well as Gordon Pask's description of how the mutual learning process between humans and computers to reach agreements through dialogue (Pangaro & Dubberly, 2014). The developed AI bot is deployed in a group messaging environment, which allows a group of designers to discuss with the AI simultaneously, which is a more realistic simulation of a real-world practice scenario.

Therefore, this paper proposes an AI-collaborated design framework and a developed agentic AI system employed in a collaborative design environment. The main body of the paper first describes the proposed design framework and the mechanism of the agentic AI system. Then, a case study with three design process examples is presented. The different observed approaches will be analysed, followed by the concluding implications in AI-collaborated design process research discourse process, envisioning potential applications in real-world practice and future research directions.

2. Methodology

Figure 1 illustrates the overall design workflow, composed of two parts: AI-collaborated performative optimisation-based design and AI-collaborated style testing. The workflow begins with performative design optimisation, in which designers produce massing models of varied building configurations and forms. Instead of selecting the best-performed option for further development, designers are encouraged to select the preferred options that can be suboptimal in performance and discuss with the agentic AI with both massing model image and text inputs. The discussion could enhance the evaluation and comparison of selected options, suggest how the optimisation could be improved, and potential styles that could be experimented with in the next stage. After styles are tested with AI art generation tools, the selected images could be sent to the AI bot for style iteration discussion. AI bot could provide verbal feedback and generate reference images as visual information for further iterations.

For an efficient experiment, the first part is demonstrated by using EvoMass on the Rhino-grasshopper platform (Wang et al., 2024), while EvoMass can be replaced with other parametric models and optimisation workflow. The second part is illustrated with Stable Diffusion, an open-sourced AI art generation tool which allows images with controlled features such as massing forms. For AI collaboration, the agentic AI is employed as a chatbot on WeChat, a common messaging application in mainland China; as such, human designers could proceed with a conversation with the AI in a group chat environment with a minimum learning curve.

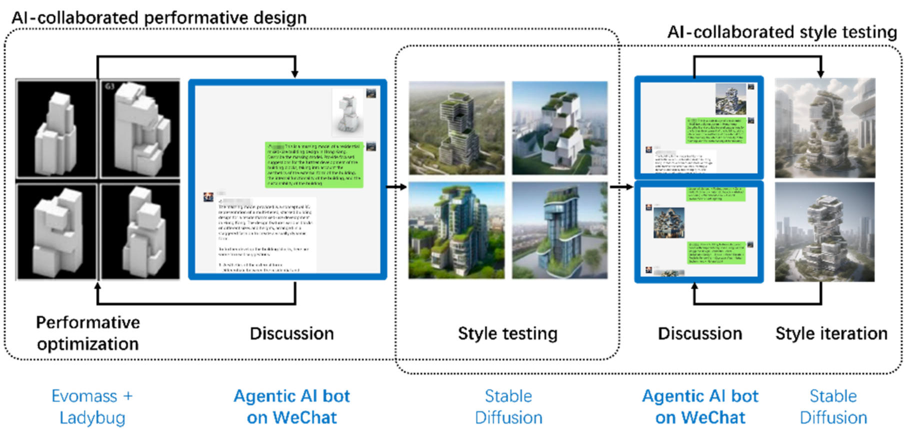

Figure 1. Proposed AI-collaborated Design Framework

2.1. AGENTIC MULTIMODAL AI SYSTEM

The agentic multimodal AI is developed as an architectural-focus AI design partner; the schematic of the proposed system is illustrated in Figure 2. After receiving inputs from the designer (texts and images), a multimodal large language model (LLM 1) acts as a question classifier to identify whether the enquiry includes massing model images or renders. If it includes a massing model image, it is assumed that the designer is working at the early stage so that the preset workflow would be most comprehensive.

It first initiates a web search through the Google engine to gather relevant case studies according to the geometrical features of the massing model. Three case studies information would be collected and abstracted, considering the balance between web searching speed and designers' cognitive load. LLM 2 has a pre-entered project brief and a knowledge base that includes the EvoMass manual and research articles about the performative architectural design process. The abstracted search results would be sent to LLM 2 as a reference, together with the knowledge base, to provide design advice in three different directions and suggest a preferred direction, followed by a prompt for the preferred option. The prompt will then be sent to the image generator to generate a reference image. Finally, the results gathered from the previous steps will be compiled as the final output. If the designer's inputs include a render, it is assumed that the design progress is more advanced so that no case studies would be searched for, but still providing different advice directions and generating a reference image according to the enquiries and preferred design direction. If the inputs either include a massing image or rendering, the enquiry would be processed by LLM 3, a multimodal LLM which is also equipped with a project brief and knowledge base, and it has the freedom to decide which tools to use, including web-search and image generation.

The whole agentic multimodal AI system is developed using the LinkAI platform (LinkAI, n.d.). In this setup, all LLMs are based on OpenAI's GPT4o, and the image generator is based on OpenAI's DALLE-3 after balancing cost, speed, functionality (vision-enabled) and effectiveness.

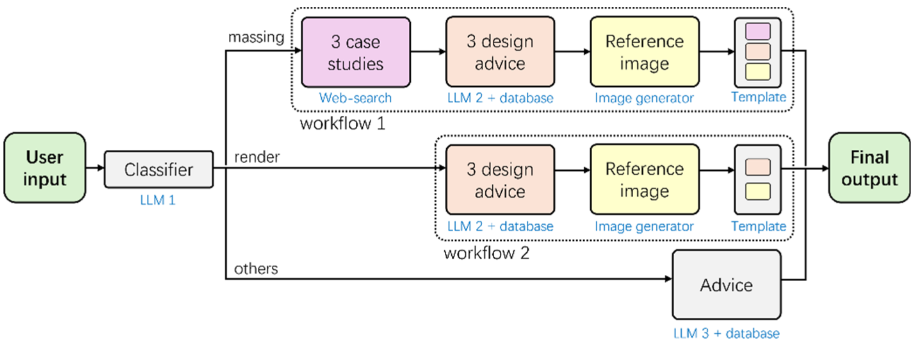

Figure 2. Schematic of the Proposed Agentic AI System

2.2. COLLABORATIVE DESIGN ENVIRONMENT INTEGRATION

The developed AI system is then employed as a chatbot on WeChat by integrating Chatgpt-on-wechat, an open-sourced project (zhayujie, 2024). It provides a seamless integration into user-friendly communication methods such as voice messages, on-device uses such as mobile phones, etc., and simulates real-world practice scenarios.

3. Demonstration

A design workshop for undergraduate architecture students was organised to demonstrate the proposed design workflow and agentic AI system. The design project was a mixed-use building cluster renovation in Gusu old town in Suzhou, China. Most Chinese old town renovations focused on façade repainting, with little consideration of architectural typology improvement or preservation of Chinese architectural features. For the neighbourhood scale, residents lack public space, and side streets are dark due to trees and adjacent buildings (Figure 3). Therefore, the students were encouraged to propose a new typology that considers urban daylight performance and local people activities while integrating traditional architectural styles.

At the beginning of the design workshop, all tutors and three students joined a WeChat group with the AI bot as a design team; the three students developed their concepts, respectively. The different AI-collaborative approaches are displayed below, and the key conversations that drove the design process are presented. At the same time, within the group chat environment, the students can observe their peers' utilisation of the AI chatbot and get inspiration for their design.

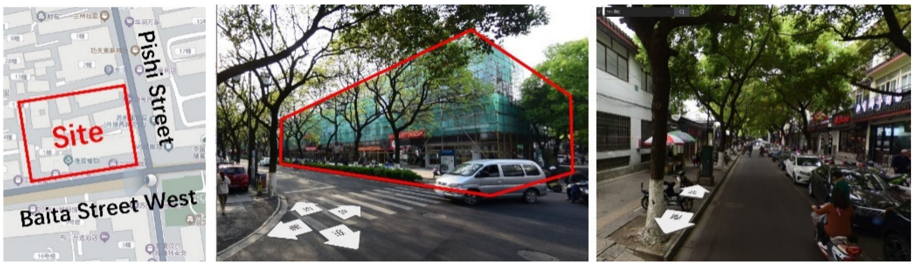

Figure 3. Site Plan and Photos of Renovation Block and Side Road, from Baidu Map

3.1. STUDENT ONE: PERFORMANCE-BASED

Student One had the design idea that the newly designed building could enhance the neighbourhood's living quality by maximising the sunlight hours in the surrounding buildings. After the first round of performative optimisation, a desired option was selected manually to initiate a design conversation with the agentic AI system, asking for further design development advice.

Since the input was neither merely a massing model image nor a render, it triggered LLM 3 of the agentic AI system. It began with explaining the simulation diagram results and providing five different design advice. The student found two pieces of advice intriguing, including "Step-like setback design reduces blockage on the south and west sides" and "using the rotation function provided by EvoMass to reduce shading effect to surrounding houses", especially the second advice since the student did not realise such function exists in EvoMass. At the same time, the AI pointed that out with relevance to the design intent of minimising shading for the adjacent houses. Then, one of the tutors supplemented a follow-up question to the AI to elaborate on how the student might use the EvoMass rotation function. Then, the student proceeded with other rounds of performative optimisation with consideration of a stepped podium for Southwest facing modules and rotation for the rest of the massing. After further manual modelling of the selected option, the massing model was sent to the AI again for further development.

Since the second round of conversation began with a massing model image, it triggered Workflow 1, which provided web-searched case studies, design advice and reference images. The student found the response comprehensive and relevant but too general. A tutor then asked a follow-up question such as "Regarding the comment on how the stepped design could enhance view while protecting the user privacy, please elaborate and provide specific and explicit design advice" to enhance communication to be conversational. It was observed that the student then initiated rounds of follow-up questions to request more variety of reference image generations and asked the AI why it made such suggestions. Finally, the student selected the desired reference image and edited the prompts using AI to have multiple style tests using Stable Diffusion.

Figure 4 summarises the design workflow. The conversation mostly surrounded the performative design process; the agentic AI system displayed the capability of finding relevance between design advice and optimisation tools. It inspired Student One to improve the performative design strategies by applying EvoMass's rotation function on the street-facing massing and a subdivision optimisation for the Southwest-facing massing. As Student One was the first student to initiate a conversation with the AI system, it was observed that the student showed confusion when the AI response was unexpected or too general since this mode of AI-collaboration is unconventional. Then, more sophisticated designers (tutors in this case) could act as a "student-assistant" by asking follow-up questions to catalyse the AI-human collaboration.

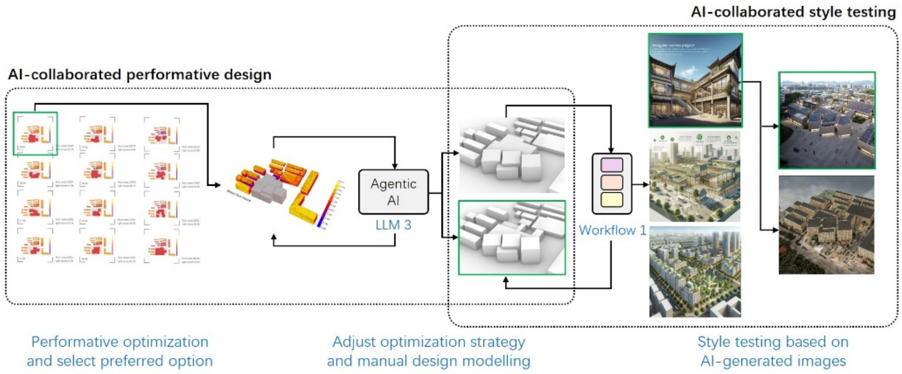

Figure 4. Design Process by Student One

3.2. STUDENT TWO: EVALUATION-BASED

Student Two began design optimisation from a more practical perspective using total floor area and sunlight hours as the optimisation objectives. Since it is observed that there was an intention for the massing to try its best to be distant from the surroundings and close to the main road, the strategy was then revised by adding an objective to provide diagonal internal roads after getting inspiration from Student Three. After the second round of optimisation, Student Two was confused about how to select the desired option to develop further. The tutors then suggested that both options might be sent to AI for giving suggestions. Since the comparison of multiple designs was not considered during the design of the agentic AI system, the enquiries triggered LLM 3 to proceed. From the lessons learnt from Student One experience, specific questions were asked, such as "Provide explicit comparison of the massing models including potential challenges and opportunities, and how to further develop in terms of massing models and style testing respectively". After several rounds of conversation, a massing model was selected, manually revised and sent to AI for advice on style testing.

It automatically triggered Workflow 1 to provide a comprehensive response. Since Student Two was unwilling to directly apply the generated image as a reference for style testing, abstracts of verbal advice were compiled into different prompts manually for the first round of style testing. With the successful conversation process of discussing the massing models with AI, Student Two then sent the different renders to AI for comparison. After rounds of conversations, instead of manually compiling prompts for the next round of style testing, Student Two asked the AI to compile prompts with specific design directions automatically. Student Two also treated the agentic AI system as a team of designers. After completing the second round of style tests, Student Two also knew more about the agentic AI mechanism. Student Two was then inspired that for the next round of style tests, not only were the three images compiled to AI for comparison by LLM 3, but also sent one at a time to trigger Workflow 2 to proceed with in-depth analysis. In the final design iteration, Student Two finally picked an AI-searched case study from a previous conversation, two AI-generated images and human-AI-compiled prompts for the final three style tests.

The AI-collaborated design process of Student Two is shown in Figure 5. Compared to other students, Student Two was more open to any design directions. Then, more all-rounded conversations with AI were observed, while AI mainly evaluated and compared different possibilities. Meanwhile, the flexibility of AI-collaborated conversation is observed. Student Two invented ways of transferring information from AI to style testing; case studies, verbal analyses, or reference images could all become sources of style tests. Throughout the process, Student Two showed that the design conversation could benefit from not aiming to find the best solutions but diverging design possibilities.

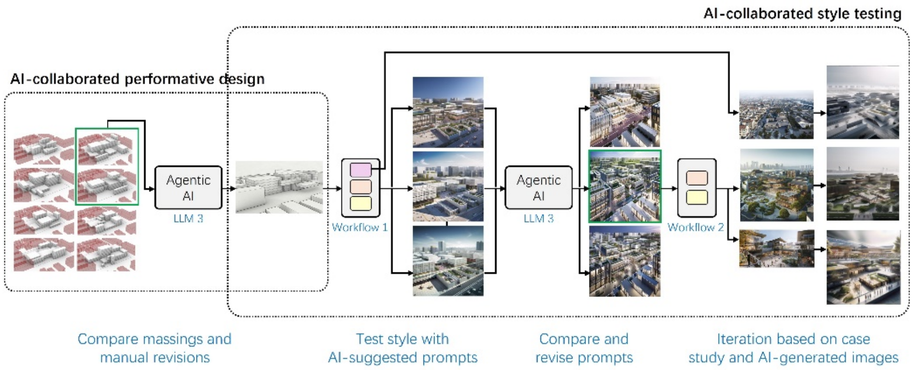

Figure 5. Design Process by Student Two

3.3. STUDENT THREE: DESIGNERLY-BASED

Student Three had a strong design idea of providing open internal streets for the local community and a deep interest in and experience with AI image generation. Two rounds of performative optimisations with direct sunlight hours, visibility and internal street openness as objectives proceeded rapidly. Then, a preferred massing model was selected and tested in various styles, using different AI models with self-designed prompts. Throughout the described process, there was no involvement of the agentic AI system and little interest in the AI experiment was shown from the beginning. Until Student Two's experience showed that AI was capable of comparing design options, Student Three attempted to ask AI to evaluate the different renders. It was observed that with other students' attempts as references, Student Three made quick adaptations by actively asking the AI for clarification and elaboration without intervention from tutors. For example, when the AI first compared the five selected renders and suggested the "best" option, Student Three showed doubt and made a follow-up question ", What if we consider local residents' living quality as the first priority?" to see if what are the "hidden" design judgement behind AI. When confusion was found in AI response, Student Three would ask focused questions such as "How would you balance 'newly designed open street life' and 'disruption of traditional lifestyles' as you suggested? Please advise explicit design methods to achieve so.". Finally, the useful comments from the AI were compiled into prompts and asked the AI to generate reference images. The images were then transferred to the design image for quick style experiment iterations.

As shown in Figure 6, designers who had strong design intent, such as Student Three, might neglect unfamiliar tools, such as AI chatbots, in the first place. However, when explicit useful examples of AI-human collaboration emerged, Student Three showed immediate adaptation in the design process. Student Three only had a conversation with LLM 3 throughout the process because the enquiries focused on very specific design aspects. This indicates that despite pre-planned workflows providing comprehensive information, a highly flexible and adaptive AI agent such as LLM 3 is necessary for in-depth conversation.

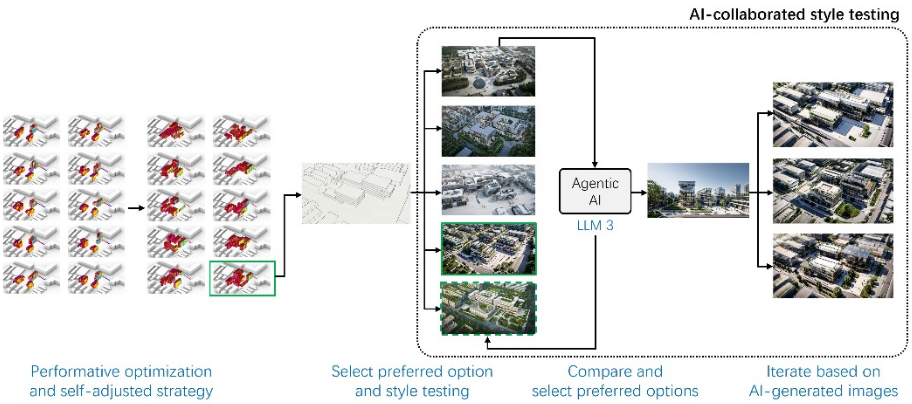

Figure 6. Design Process by Student Three

4. Discussion and Conclusion

The three demonstrations display the feasibility of how an agentic multimodal AI system could be integrated into a collaborative design environment during early-stage architectural design exploration. The proposed framework and developed AI tool address three common issues in today's AI-collaborated design processes: one-to-one communication, the high learning curve of using sophisticated AI systems, and a low understanding of architectural discipline during conversation. Three examples are demonstrated by three undergraduate architecture students. They displayed three different approaches: performance-based, evaluation-based, and designerly-based. The variety of approaches illustrate how the proposed framework and AI system enable the design process beyond master-slave relationships, as described by Negroponte (1976), and enhance the iteratively evolving, conversational design process, as envisioned by Schön (2017) and Pask (Pangaro & Dubberly, 2014).

4.1. IMPLICATIONS TO RESEARCH, AI IN EDUCATION AND PRACTICE

From the perspective of research and AI education in architectural design, the proposed framework illustrates the minimum learning curves of AI tools. It allows students to dedicate themselves to the design process with little time investment in learning advanced AI tools. The designed workflows show potential to students, especially those with little experience, who tend to ask fewer comprehensive questions for design advice. The more flexible agent (LLM 3) allows more students to ask unexpected and follow-up questions. Collaborative environment integration allows tutors to participate in the conversation by asking follow-up questions. Also, mutual-learning conversation techniques and new AI-collaborated design methods are possible between students, according to the observation presented in the workshop. Regarding real-world practice, it demonstrates the potential to extend the collaborative scenario with other stakeholders such as engineering consultants, clients etc. Project information and domain knowledge, specific from different expertise, are foreseen to be integrated with the AI agents by embedding a knowledge base system.

4.2. LIMITATIONS AND FUTURE RESEARCH

There are several challenges and limitations that suggest potential research directions for the developed agentic AI system design, which currently includes only two workflows and one agentic AI. A more generalised design approach is necessary for better adaptability to different designers. Although the case study was performed in a collaborative environment, students primarily developed the projects independently; further experimentation with multiple designers on a single project could test the effectiveness of the proposed framework.

Regarding the conversation with AI, most students found AI responses to be general until prompted with follow-up questions. Despite using WeChat as the medium of AI-human collaboration, which enables immediate use, the switch between AI conversation and other AI tools, such as Stable Diffusion, still poses an apparent learning curve to students with no prior AI application experience. Additionally, responses often lack connections between abstract concepts, such as connections between design software function and design ideas, suggesting the need for advanced techniques like multi-agent systems and knowledge graphs. Finally, this study was conducted with a limited focus group of three participants. Future research will expand data collection to encompass a larger and more diverse participant pool to ensure a comprehensive understanding of user behaviour.

4.3. CONCLUSIONS

To conclude, this paper proposes an AI-enhanced design framework for a collaborative design environment. The design framework is supported by a developed architectural-focus agentic multimodal AI system integrated into WeChat, a common messaging application allowing group discussion. The feasibility is demonstrated through a design workshop. Case studies of three students' AI-collaborated design process show a variety of design approaches for early-stage architectural design applications. Finally, the potential implications of the research are discussed, and challenges and future research opportunities are identified.

Acknowledgements

This research is supported by the Postgraduate Research Scholarship (PGRSB211206) and Research Development Fund (RDF-23-01-107) offered by Xi'an Jiaotong-Liverpool University. The authors declare that the research compiles with relevant ethical regulations, with approval from Xi'an Jiaotong-Liverpool University Research Ethics Review Panel. The authors would like to thank the students who contributed to the design examples presented in this paper: Sizhe Chen, Yibing Shi and Guanhua Bao for the first, second and third examples, respectively.

References

Abedin, B., Meske, C., Junglas, I., Rabhi, F., & Motahari-Nezhad, H. R. (2022). Designing and Managing Human-AI Interactions. Information Systems Frontiers, 24(3), 691–697. https://doi.org/10.1007/s10796-022-10313-1

Bolojan, D., Yousif, S., & Vermisso, E. (2024). Latent Design Spaces: Interconnected Deep Learning Models for Expanding the Architectural Search Space. In M. Barberio, M. Colella, A. Figliola, & A. Battisti (Eds.), Architecture and Design for Industry 4.0: Theory and Practice (pp. 201–223). Springer International Publishing. https://doi.org/10.1007/978-3-031-36922-3_13

Ibarrola, F., Lawton, T., & Grace, K. (2024). A Collaborative, Interactive and Context-Aware Drawing Agent for Co-Creative Design. IEEE Transactions on Visualization and Computer Graphics, 30(8), 5525–5537. https://doi.org/10.1109/TVCG.2023.3293853

LinkAI. (n.d.). LinkAI. Retrieved 4 October 2024, from https://link-ai.tech/portal

Negroponte, N. (1976). Soft Architecture Machines. MIT Press, MA. https://doi.org/10.7551/mitpress/6317.001.0001

Pangaro, P., & Dubberly, H. (2014). What is Conversation? How Can We Design for effective Conversation? In M. Shamiyeh (Ed.), Driving Desired Futures: Turning Design Thinking into Real Innovation (pp. 144–159). Birkhäuser. https://doi.org/10.1515/9783038212843.144

Schön, D. A. (2017). The Reflective Practitioner: How Professionals Think in Action. Routledge. https://doi.org/10.4324/9781315237473

Wang, L., Janssen, P., & Ji, G. (2024). Optimization-based design exploration of building massing typologies—EvoMass and a typology-oriented computational design optimization method for early-stage performance-based building massing design. Frontiers of Architectural Research, 13(6), 1400–1422. https://doi.org/10.1016/j.foar.2024.06.001

zhayujie. (2024). Chatgpt-on-wechat [Python]. https://github.com/zhayujie/chatgpt-on-wechat (Original work published 2022)

Zhong, C., Shi, Y., Cheung, L. H. & Wang, L. (2024). AI-Enhanced Performative Building Design Optimization and Exploration: A design framework combining computational design optimization and generative AI. Proceedings of the 29th CAADRIA Conference, Singapore, 20-26 April 2024, Volume 1, Pp. 59–68. https://10.52842/conf.caadria.2024.1.059