Proceedings of the Design Society, Volume 5: ICED25
https://doi.org/10.1017/pds.2025.10264

Multimodal generative AI for conceptual design: enabling text-based and sketch-based human-AI conversations

Gaelle Baudoux✉, Chenjun Guo and Kosa Goucher-Lambert

University of California, Berkeley, USA

✉ gbaudoux@berkeley.edu

ABSTRACT: Recent advances in AI offer promising opportunities for creative design, particularly through the generation of inspirational images. While prior research has explored the general benefits and limitations of text-to-image tools, there is significant potential in overcoming these constraints by investigating agile, multimodal prompting to facilitate more project-appropriate human-AI interaction. We present the development of a system designed to support both text-based and sketch-based image generation, serving as a research artefact for studying creativity support through multimodal Generative AI. The system enables dynamic dialogue interaction and visualization of the respective contributions. This paper focuses on the development of this AI system as a research artefact to enable future research through design, exploring how multimodal prompting can influence the design process.

KEYWORDS: conceptual design, creativity, computer aided design (CAD), collaborative design, artificial intelligence

1. Introduction

The early stages of design, particularly ideation, are critical in determining the performance of the final artefact. This ideation activity can be stimulated and enhanced through various methods (Casakin & Wodehouse, 2021). One promising method is co-design with AI tools, such as the use of generative AI systems, which stimulate creativity and help overcome cognitive fixation through rapid and expensive design exploration (Karimi et al., 2020; Kim, Maher, & Siddiqui, 2021; Enjellina, 2023). Generative AI tools, such as the well-known Midjourney or DALL-E, whose use exploded in 2022 (Enjellina et al., 2023), are generative software that produce collage images of high aesthetic quality based on textual or, more recently, sketched parameters, entered by the user in a so-called prompt (Enjellina et al., 2023). Recent research explores the general benefits and pitfalls of generative AI systems for design (Enjellina et al., 2023; Beyan & Rossy, 2023). However, the design of agile, non-disruptive co-design systems remains a significant challenge (Rezwana, 2023). Specifically, how different prompting modalities can affect the design process and design outcomes is not yet fully understood. The overarching objective of our research is to understand "*how multimodal prompting mixing text and sketch inputs, can better match naturalistic ideation activities and positively impact design processes and outcomes*".

To investigate this objective, our research utilizes a research through design (RtD) approach. This research approach is primarily driven by the fact that no current GenAI-enabled design tool currently exists that affords the features we require to investigate our research objective. As such, we primarily contribute a custom multimodal generative AI system that supports both text and sketch prompting in different generation modes - from convergent to divergent, and from concrete to abstract. Our system serves as a research artefact to expose designers to this multimodal prompting feature, and to collect data on human-generative AI conversations and their impact on design activity.

The present paper summarises the current state of text-based and sketch-based image generation tools and explains why we needed to develop a new prototype to investigate multimodal generative AI. It then identifies promising human-AI strategies from the literature that can be leveraged in the development of AI co-design tools. The paper details the prototype developed as a research tool and evaluates its functionality to support the research investigation with a pilot user study.

2. Related work

This section analyzes recent applications of generative AI for design ideation, including key related work on text-based and sketch-based image generation, and introduces key concepts regarding human-AI co-design.

2.1. Generative AI for creative conceptual design

A literature review of recent - since 2019 - applications of Generative AI tools within the context of design ideation activities revealed several researches presenting Generative AI tools application among various subfields of design such as graphic interface design, product design, creative activities, engineering, and architectural design (N=46 papers). Ultimately, 26 papers employed text-based and sketch-based AI image generation to stimulate creative, engineering, or architectural design activities. Regarding text-to-image generation, studies highlighted that text-based image generation can support architects during the early stages of design (Paananen, Oppenlaender & Visuri, 2024), particularly in open-ended concept ideation (Nagele, 2023). These generative AI tools offer the potential for rapid, expansive design exploration (Enjellina et al., 2023). Furthermore, Beyan and Rossy (2023) show that generative AI tools can facilitate both abstract thought and the production of tangible design outcomes. Casakin and Wodehouse (2021) also demonstrated that by transcending the limitations of realism and physical constraints, such systems enable designers to push the boundaries of imagination and explore unconventional concepts. However, these studies also highlight several limitations of text-to-image generation tools, including their inability to address specific design goals (Nagele, 2023) and a tendency to produce outputs that are overly reductionist or unrealistic. Zhou et al. (2024) argue that text-to-image models rely on a recognition-based process mediated by natural language, whereas traditional art and design often involve direct manipulation of visual elements, such as color and shape. This fundamental difference, Zhou et al. (2024) suggests, restricts the creative freedom users experience when working with text-to-image generation tools.

Regarding sketch-to-image generation, Zhang et al. (2023)'s findings suggest that most designers believe AI can inspire creativity and enhance design sketching. However, Zhang notes that general sketch-to-image generation tools accessible to the public lack an understanding of design knowledge, requiring significant effort to adjust parameters to achieve the desired results. To solve this problem, Gao (2024) developed a domain-specific urban design sketching platform that incorporates urban design knowledge, with intuitive sketches as symbolic inputs for generating urban design outputs.

Finally, a couple studies have explored combining multiple modalities in generative tools. For instance, Kwon et al. (2022) built a multi-modal platform to retrieve 3D-model parts based on similarities in visual and functional features to 3D-modeled inputs specified by the designer. Or the Sketch2Prototype model (Huang et al., 2022), that processes hand-drawn sketches through sequential stages: sketch-to-text, text-to-image, and image-to-3D, ultimately converting sketches into 3D models. The model also allows users to modify the text generated during the sketch-to-text stage to improve the accuracy of the final output. However, this model does not thoroughly observe and examine user behavior, particularly the differences between using text input and sketch input.

This background synthesis reveals that while current generative AI tools excel in their specific tasks, they lack alignment with the characteristics of naturalistic ideation sketches. Among the few existing multimodal generative AI tools, none adequately address our research question, highlighting the need for a new system to be developed to investigate how to better support naturalistic design ideation through multimodal input.

2.2. Human-AI collaborative ideation and co-design

In contrast to traditional image generation systems, co-design agents, defined as an AI collaborating alongside a human designer into a unified process where their individual roles become indistinguishable (Liapis et al., 2016), require bidirectional information exchange between the designer and the AI

(Rezwana, 2023), and effective coordination and communication are essential for successful collaboration (Seeber et al., 2020). Two key factors facilitate this collaboration: first, the alignment of AI agents with human cognitive processes, making them more akin to a person's mental system than traditional tools (Stoimenova & Price, 2020); and second, the shift from a hierarchical tool-user relationship to a collaborative, partnership-based dynamic (Seeber et al., 2020; Figoli et al., 2022). AI agents are capable of inductive-deductive behaviors, including the inspiration and evaluation of design solutions (Figoli et al., 2022), but they must be adapted to human design strategies (Rezwana, 2023). In their synthesis of the human and generative AI workflows, Enjellina et al. (2023) conclude that the human brain processes mental images in order to create images based on emotional responses and memories of past experiences, process that is the key value of inspirational stimuli during design (Hu, McComb & Goucher-Lambert, 2023). Meanwhile, the AI system requires human input to generate and combine images. In this manner, humans act as operators who create and operate AI systems as a tool for the retrieval of inspiration. This shift in the Human-AI relationship from AI as a tool to AI as a partner also alters the role of the designer, who transitions from task execution to evaluating and making decisions about AI-generated ideas (Figoli et al., 2022).

This subsection highlights the shift from a traditional tool-user relationship to a more collaborative, partnership-based interaction, where AI acts as a co-design partner, supporting the designer with inspiration, evaluation, and iterative development. Our work, by focusing on adapting AI tools to human design strategies through multimodality bidirectional exchanges, contributes to refining AI systems that enhance co-design practices and foster agile human-AI teamwork.

3. Multimodal Gen AI prototype system development

To support our research, questioning "*How prompting modalities, and in particular multimodal prompting that mixes text and sketch inputs, can better match naturalistic ideation activities and positively impact design processes and outcomes*", we develop a design tool that supports multimodal human-AI co-design in a research through design approach. The following subsections present, respectively, the strategies derived from related works that shaped our design decisions, the general principles of the system, and its detailed architecture. Then, using the developed system we conduct a series of preliminary user tests to evaluate its functionality in investigating design activities.

3.1. Human-AI interaction-informed system design strategies

The related work section highlighted that our system should be supporting bidirectional information exchange, designed to follow along with the user, and both foster engagement while providing goal-oriented contribution to the design. To overcome the current Gen AI tool's limitation we should develop a system that incorporates architectural knowledge to be able to address specific design goals and produce realistic outputs.

Research on Human-AI collaboration emphasizes that to interact with AI as a true collaborator, the AI should outperform the human agent in specific tasks. This helps avoid cognitive overload, where the designer must continuously adjust or exclude AI contributions (Figoli, Mattioli, & Rampino, 2022). However, Figoli and colleagues show that, when AI is used as an external stimulus (as in our case), this rule is not critical: the role of AI depends more on the design configuration - either continuous collaboration, where AI leads the creative process, or alternating collaboration, where AI assists a human-driven process. Zhang et al. (2021) also observe that AI boosts low-performance designers but can reduce the performance of high performers due to cognitive overload. Therefore, AI systems should offer straightforward, digestible outputs in small quantities, keeping users engaged without overwhelming them. For our use case, this suggests that users don't require the AI to outperform them, but they must remain engaged with it to enhance creativity. Additionally, AI should only intervene when needed to avoid cognitive overload and maintain an alternating collaboration configuration that supports a human-driven creative process.

Specifically in human-GenAI collaboration, the researchers analyzing current Generative AI tools' use state that its first benefit is to help designers translate abstract ideas into tangible design outcomes (Beyan & Rossy, 2023). However, the literature identifies limitations: the need for users to stop their workflow to prompt image generation, the challenge of prompt engineering for accurate results, and the lack of ability to edit generated images (Nagele, 2023; Enjellina et al., 2023; Beyan & Rossy, 2023; Zhang et al., 2023;

Paananen et al., 2024). Thus designers need more accurate images they can better control the generation as well as an agile way of interacting with our AI system.
Specifying the collaboration style, the participation of our AI agent can be either in parallel to the human agent's one or in turn-taking. We choose the latter to give the human the decision-making power, to enhance their engagement in the interaction and limit the overload for the high-performance designers. The task can be divided between the agents or the same, and in our case it will be the same to explore collaboration activities and avoid cooperation activities. Finally, the timing of the AI agent's input can be either spontaneous or planned. As we want to study multiple conversation modalities, the AI intervention has to be planned. On the other hand, specifying the communication style, the human to AI communication can be by voice, direct manipulation, embodied or by text. We drop the first one that is unrealistic in professional working environments but keep all the remaining to give the freedom to the user and ensure we provide the one they will feel most engaging. The AI to human communication can be by speech, text, embodied, haptic or visual. As AI generated images show undeniable benefits directly linked with the visual nature of the output (Casakin & Wodehouse, 2021; Enjellina et al., 2023; Beyan & Rossy, 2023; Paananen, Oppenlaender, & Visuri, 2024), we will enhance this aspect by having visual outputs for the AI contribution.

3.2. General principle of the system developed

Our system is designed to facilitate ideation and design tasks, enabling users to develop their concepts through sketching and collaborative interaction with the AI. The human-to-AI communication is facilitated by two modes of conversation: text-prompting and sketch-prompting, which includes additional parameter specification. Conversely, the AI-to-human dialogue employs two modes of image generation: rendering with high fidelity and inspiring which incorporates a chosen reference style and allows for greater divergence from the prompt. Subsequently, three generated images are displayed. The human operator may elect to discard some of the generated images or add them to the project library. Furthermore, the designer may opt to incorporate some of the AI's suggestions into their design and sketches. We chose to incorporate text and sketch prompting because these two modalities are the communication modalities intrinsically present in the naturalistic design sketches. Furthermore, we chose to incorporate both convergent rendering generation and divergent inspiring generation as these two types of visuals had been identified as both needed and powerful to aid the variety of idea generation behaviors in a previous study (Baudoux & Safin, 2025).

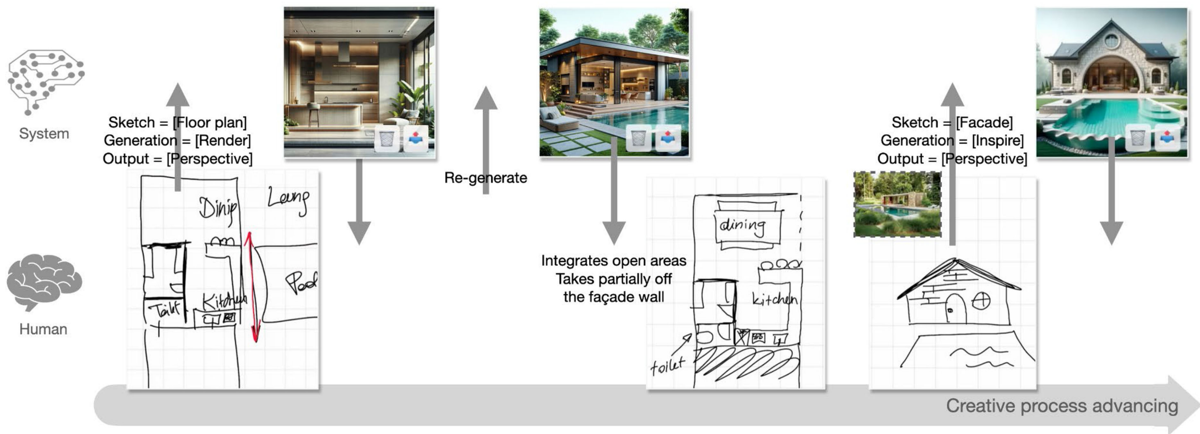

Figure 1. Extract of conversational co-creative loops between a designer and our AI system

Figure 1 shows a sketch mode conversation with the AI system, where the user co-designs a pool house for a mansion's backyard. The user starts by expanding on initial ideas, resulting in the AI suggesting a covered outdoor area, an idea integrated by the user for the dining area. The user then requests an alternative, more traditional, style, guiding the AI with a reference image.

3.3. Detailed system architecture

We developed the **interface** to support both activities of design (sketching, evaluating, project synthesising), activities of communication with the AI (text prompting, sketch prompting, generated images visualization and evaluation), and data collection by triggering explicit interface's actions recorded in the back-end (generating button, trashing image button, adding to project mood board button). This interface is designed to firstly ensure that the AI system could easily access the live sketch and secondly to smooth the user experience load by allowing them to use only one interface. The user can start conceptualizing their ideas of design by sketching on the sketching space of the homepage (Fig. 2 - A). When wanting to interact with the AI-partner, the designer can select the desired mode of conversation: text (Fig. 2 - B) or sketch dialog (Fig. 2 - C), depending on their needs at the moment, the preciseness of their idea, etc. For the sketch conversation mode, the designer can specify additional parameters, such as the type of item sketched and the type of output image wanted (Fig. 2 - D), to help the system understand the sketch and the desired contribution. The designer can also specify the desired generation mode expected from the AI: precisely rendering the prompt (Fig. 2 - E) or providing inspiration with more divergent propositions fitting a specific reference style image (Fig. 2 - F). Once generated by the AI, three images are displayed to the designer (Fig. 2 - G) who can choose to discard it, sending the signal that the AI wrongly understood what they prompted (Fig. 2 - H), on the other hand choose to add some to the project library (Fig. 2 - I), meaning that the AI contribution is interesting enough to be added to the project, or in between do nothing in particular, and the image will stay in the image library (Fig. 2 - J) for further consultation if wanted.

Figure 2. Software architecture and interface visuals

Figure 3 and Table 1 presents the details of the **prompt structure** to move from the user's 2 by 2 main modes of dialog (i.e. sketch/text X render/inspire) to the received AI response with a set of three generated images.

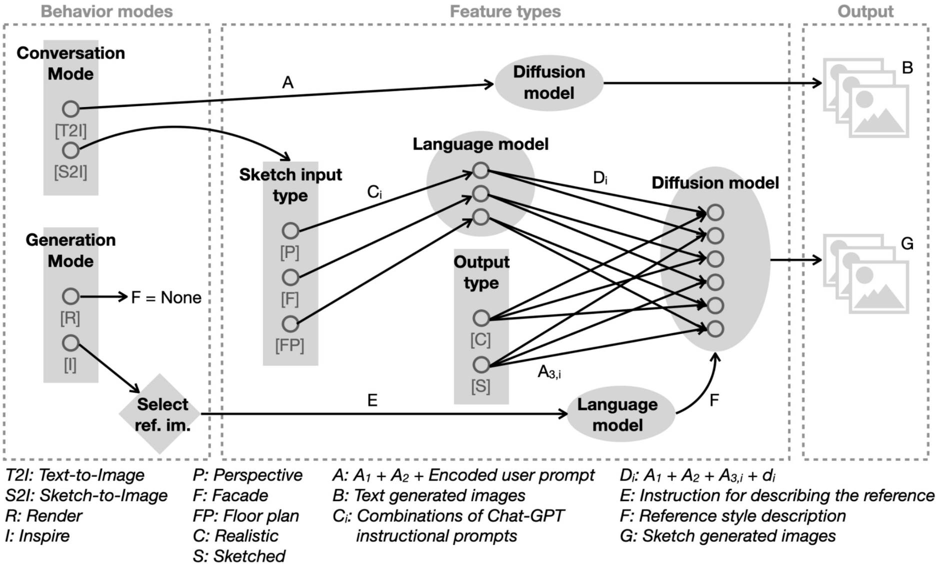

Figure 3. Diagram of the prompt structure

The **targeted design problem** addressed specifically in the case study of development of this new tool is to design the architecture of an accessory dwelling unit. The system is thus calibrated, in terms of prompting fixed parameters and in the training of the Generative AI model selected to be the most performant on this type of design task.

Table 1. Prompt's textual fixed parts

A_1	Generate a professional rendering of the architectural design of a separate dwelling in-law suite in the backyard from the interior.
A_2	It should be a RAW photo, architecture photography, and hyper realistic. Incorporate this description:
$C_{P\text{-}P}$	This is a sketch of interior and your description will be used for creating render image, describe the arrangement of the rooms, furnitures, describe the view angle, detailed words for accurate recreation. Keep the parameters under 250 characters. Do not mention the sketch, use 'the render image contains...'
$C_{P\text{-}FP}$	This is a sketch of interior and your description will be used for creating floor plan. Describe the floor plan, recognise the largest room in the center and use this room as the start point, describe all surrounding spaces in the floor plan. Do not include descriptive language or interpretive phrases. Pay attention to the shape of room. Use only relative positioning terms like 'left,' 'right,' 'top,' and 'bottom.' Focus on relative sizes of each space in relation to the central room within one paragraph. do not mention 'floor plan'. Start with the 'the top-down view rendering contains'
$C_{FP\text{-}FP}$	This is a sketch of a floor plan and your description will be used for creating top-down view rendering. recognise the largest room in the center and use this room as the start point, describe all surrounding spaces in the floor plan. Do not include descriptive language or interpretive phrases. Pay attention to the shape of room. Focus on relative sizes of each space in relation to the central room within one paragraph. Do not mention floor plan, start with 'this top-down view rendering include'
$C_{FP\text{-}P}$	This is a sketch of interior floor plan and your description will be used for creating perspective rendering. Describe the one level interior from eye level and from one of the indoor room, recognise the largest indoor room in the center and use this room as the start point, describe all surrounding spaces based on how they would appear in a perspective view. Do not include descriptive language or interpretive phrases. Pay attention to the shape of room. Use only relative positioning terms like 'left,' 'right,' 'front,' and 'behind.' Do not use 'above' 'below' or any description that implies a top-down view. Observe the exact spatial arrangement as shown in the image. Avoid terms like 'floor plan,' 'layout,'. Focus solely on the spatial arrangement and relative sizes of each space in relation to the central room within one paragraph. Start with: the interior contains
C_F	The attached image is a sketch of a separate dwelling in-law suite facade in the backyard. Analyze this architectural sketch in precise, detailed words for accurate recreation. Keep the parameters under 250 characters. Do not mention the sketch, use 'the render image contains...'

(*Continued*)

Table 1. Continued.

d_P	The rendering should be a small separate dwelling in-law suite in the backyard
d_{FP}	The rendering should be a small separate dwelling in-law suite in the backyard. Show entire room from top view. The viewer should be right on top and in the center of the floor. It is only for one room.
$d_{F\text{-}FP}$	The rendering should be a small separate dwelling in-law suite in the backyard. Show entire floor from top view. The viewer should be right on top and in the center of the floor.
$d_{F\text{-}P}$	The rendering should be a small separate dwelling in-law suite in the backyard. Show entire floor from top view. The viewer should be right on top and in the center of the floor.
E	Write a description about this image in one paragraph. Include information on the color palette, mood board, furnishing style, material, architectural style and lighting, do not include detailed furniture, emphasize on the most important characteristic

4. Pilot user study

4.1. Task and population

To assess the functionality of our system, we ask three users from the design domain to explore ideas collaboratively with the AI. We ask them to design the highest number of new ideas for an in-law suite in a separate dwelling with the proposed AI partner within 10 minutes per prompting modality (20 minutes total). The three preliminary users have different levels of familiarity with GenAI tools - not familiar, mildly familiar and heavily familiar, are male, female, and non-binary, and were randomly assigned to start with each of the two modalities. We use Cherry and Latulipe (2014)'s Creativity Support Index as an evaluation framework. It includes a rating of 10 statements along 5 sub-axes of human-AI collaboration, answered on a scale from "highly disagree" (1) to "highly agree" (10). We administered it to the three users from the design domain after they completed the design task.

4.2. Insights from pilot study

For each test user, the task was completed successfully. In their evaluation, all three preliminary user testers were very satisfied after their trials, and all three preferred the sketch mode to the text mode, pointing out that it was easier to use and allowed them to be more creative. It is interesting to see that elaborating the text prompts was faster, but less specific: users generated 4-5 ideas during the 10-minute text prompt and 3-4 ideas during the 10-minute sketch prompt. As we can see in Figure 4, their impressions and feelings about the co-creation activity were positive. It took each of them a few minutes to familiarize themselves with the system, but then the system followed the expected mental model of behavior that they had.

In terms of user experience, the benefits identified by early test users are that the AI system actually supports and enhances creativity by generating complementary ideas and visuals. It is also a quick and easy means to achieve the generation and production of new ideas. Moreover, this system keeps the designer engaged both by maintaining their motivation, thanks to the generative aspect of the system thus acting as a partner actively proposing ideas and relaunching the activity of designing, and by allowing them to continue sketching while the AI is busy generating images, thanks to the interface design. These applications demonstrate that proposing a system that supports different modes of communication is a promising path to address a wider range of needs. Users reported that sketching was a more comfortable mode of communication with the AI when the idea was not fully formed, while writing specific keywords was a more appropriate approach when the idea was more concrete. In terms of performance, the system demonstrated the ability to adhere to the intricacies of the sketched designs. In addition, the generated images were found to be aesthetically pleasing and plausible. Within its generative characteristics, the system was effective in supporting the designer's ability to project the proposed solution while not imposing undue constraints on the designer in terms of realism.

The main limitation of the system, as most generative systems, is its dependence on the databases used to train the model used. Despite the selection of the most powerful generative models for the architectural domain, the results of the AI partner are still influenced by biases inherent in the dataset. In addition, the need to set numerous parameters when engaging with the AI partner in sketch mode may impose a cognitive load on the user. However, preliminary test users indicated that the potential for this additional effort was justified by the outcome.

The benefits and limitations observed in this pilot user test allow us to validate the interface, as it successfully supports various prompting and generation modes in an intuitive manner. Additionally, it enables users to continue sketching and designing while images are being generated. However, we identified that the underlying generative model could benefit from further training to better understand the intent behind prompts, without relying on cognitively costly specifications or presenting biases.

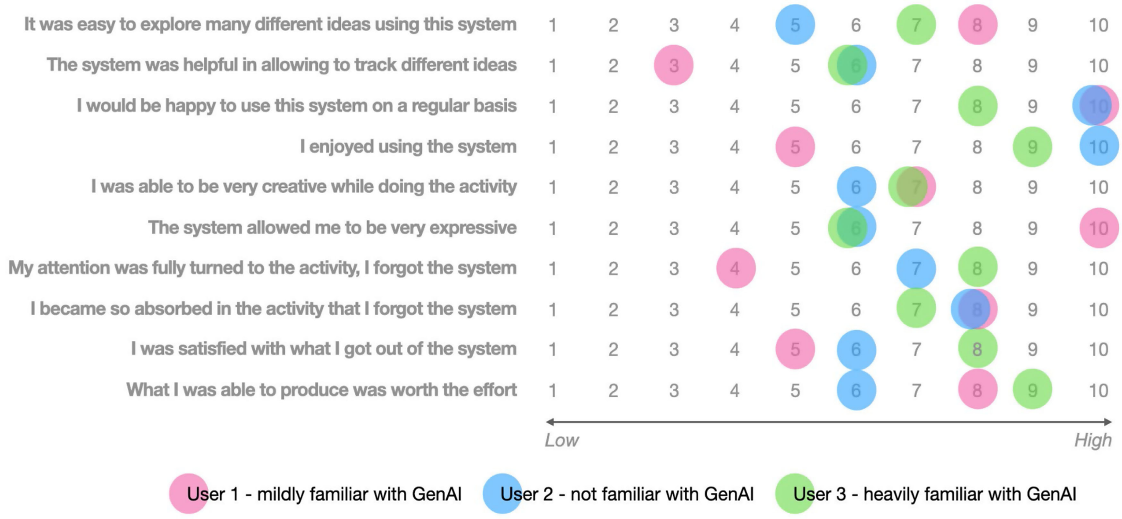

Figure 4. User tester scoring on the Creativity Support Index evaluation scale

5. Discussion

5.1. Reflection on the proposed human-AI collaboration

We consider the resulting teaming driven by the design of this new system in light of the human-AI interaction framework proposed by McComb et al. (2023), whom formalize human-AI interaction types in a 2x2 matrix - with AI being reactive (user-initiated) or proactive (taking actions without specific user prompting), and focused (specific task) or process-oriented (crossing problem boundaries). We observe that the developed system falls in the category of AI-as-tool, as opposed to AI-as-analyst, -partner, or -guide. This is due to the fact that it is reactive to human prompting rather than autonomous, with a focus on problem solving rather than across problems. In this regard, the authors note that AI-as-a-tool improves performance on key performance indicators and is a necessary position for complex problem solving, as in our case. This is an effect we indeed observed in the user testers feedback. However, by redirecting the human contribution to higher-value work, it may affect the agility of users, and this is a benefit the users stated as well. The tool developed responds promisingly to the needs identified by previous research on the subject, such as the freedom to call or not the AI partner, the need to be able to communicate vague ideas and receive in return tangible concrete results, more precise and better controlled, but in an agile way (Figoli, Mattioli, & Rampino, 2022; Beyan and Rossy, 2023). The designers were not deprived of their generative role, a risk illustrated by Figoli, Mattioli, and Rampino (2022), but shared it with the AI partner, while still retaining control over the final choices of idea implementation. The resulting instrumented process was still human-driven while being AI-augmented. Indeed, the system allows for collaborative creation, as the AI follows a see-transform-see loop: it sees the designer's text or sketch prompt, interprets it, and then transforms it into a visual, before the designer goes through the same loop of seeing the AI's suggestion, transforming his design, and rediscovering it.

5.2. Good practices for agile co-design Gen AI

Based on the human-AI collaboration literature and on the pilot user experience, we propose several recommendations for designing human-AI co-design systems. First, it is crucial to offer flexible communication modes, allowing designers to switch between sketching, text input, and direct manipulation, depending on their design stage. Sketching works well for abstract ideas, while text input is more effective for refining specific concepts. This flexibility ensures that users can engage with the AI in a way that complements their creative flow. Minimizing disruptions to the user workflow is also critical. Designers should be able to continue sketching while the AI generates images in the background. This asynchronous interaction allows for continuous engagement with the design task, avoiding unnecessary interruptions. Another key recommendation is to maintain user control over the design process. Implementing an alternating collaboration model, where the AI intervenes only when needed, minimizes cognitive overload and keeps designers in charge. The AI should enhance the design without taking over the creative process, ensuring that the user remains the primary decision-maker. Designers should also have the ability to adjust or modify the images produced by the AI, ensuring the results align with their vision and ensuring a flexible tool that adapts to the designer's needs. AI-generated outputs should inspire creativity rather than be overly realistic renderings. Designers value outputs that are aesthetically appealing and don't restrict their creative freedom. By keeping AI contributions abstract and imaginative, designers tend to build upon them rather than being constrained by rigid designs. Finally, reducing cognitive load is important for sustaining engagement. The system should simplify complex tasks and focus on providing intuitive interfaces that minimize unnecessary cognitive effort.

5.3. Support for future inspiration search modalities analysis

The developed system is intended as a simulation tool to instrument the study of co-design processes and human-AI conversational behaviors. To meet the needs of our research question, the interface supports idea visualization and human-AI communication, while logging all actions and results to serve as a research data collection tool. This successfully automatically collected research data such as each instance of the AI-provided image, sketch prompt, text prompt, image trashing action, or image import to project action and its timecode. Coupled with camera recording and thinking aloud data, we can successfully access the designers' thoughts along the process, their reasoning, and their behaviors. This will allow us to collect the necessary material to answer our research question about prompting modalities that better match naturalistic ideation activities and positively impact design. As well as specific sub-research questions such as "What is the impact of conversation modality on inspiration seeking behavior? What is the rationale behind image evaluation and selection, and is it modality dependent? Analyzing co-creation behaviors in this system makes it possible to study the idea generation pathways using AI or using each of the possible modalities, as well as the moments and frequency of these behaviors, to observe the progression of an idea over its lifetime, noting instances where it is supported by analogy with the AI contribution, or to see which AI contributions are discarded or, conversely, added to the project mood board, at what rate, and with what rationale.
Future design modification of the tool that would improve it as a research artifact will be twofold: first, overcome the limitations pointed out by the pilot user testers, and second, incorporate a built-in survey allowing the user to connect with a pseudonym, complete a pre-experiment survey with level of Gen AI background, expectations regarding the tool, demographics, etc., and a post-experience survey with evaluation of the different features proposed and self-assessment on creativity, user experience, . . .

6. Conclusion

The work presented in this paper provides insight into enhancing human-AI co-creation processes through multimodal Generative AI. Through a review of existing generative AI tools and human-AI collaboration frameworks, we highlighted that no current GenAI-enabled design tool supports multimodal prompting, combining text and sketch inputs. We also identified key strategies for designing a human-Gen AI co-design tool to address this gap. The new multimodal generative platform, developed as a research-through-design artifact, integrates both text and sketch prompting, offering distinct generation modes that facilitate divergent and convergent ideation. Preliminary user tests showed that the system effectively engages designers, fostering creativity, with users expressing a preference for sketch-based interaction due to its flexibility and alignment with natural workflows. However, users also noted limitations, such as dataset biases and cognitive load from adjusting parameters in sketch mode. While improvements are needed, the pilot tests reveal valuable opportunities for further research using this platform. Future research will focus on how these interactions can help users search for inspirational stimuli and improve ideation, as well as explore conversational behaviors and search modalities for retrieving such stimuli.

References

Baudoux, G., & Safin, S. (2025). Study of computer multi-instrumented reflexive conversation activity in preliminary architectural design. *International Journal of Architectural Computing*, 14780771241310207.

Beyan, E. V. P., & Rossy, A. G. C. (2023). A Review of AI Image Generator: Influences, Challenges, and Future Prospects for Architectural Field. *Journal of Artificial Intelligence in Architecture*, 2(1), 53-65.

Casakin H and Wodehouse A. A systematic review of design creativity in the architectural design studio. *Buildings* 2021; 11(1): 31.

Cherry, E.C., Latulipe, C.: Quantifying the creativity support of digital tools through the creativity support index. *ACM Trans. Comput. Hum. Interact.* 21, 1–25 (2014).

Enjellina, B. E. V. P., & Rossy, A. G. C. (2023). A review of AI image generator: Influences, challenges, and future prospects for architectural field. *Journal of Artificial Intelligence in Architecture*, 2(1), 53-65.

Figoli, F. A., Rampino, L., & Mattioli, F. (2022). AI in design idea development: A workshop on creativity and human-AI collaboration. *PROCEEDINGS OF DRS*, 1-17.

Gao, J. (2024). From Sketch to Design: A Cross-scale Workflow for Procedural Generative Urban Design. *Proceedings of the 29th CAADRIA Conference*, Singapore, 20-26 April 2024, Volume 1, Pp. 343–352.

Hu, M., McComb, C., & Goucher- Lambert, K. (2023). Uncovering hidden patterns of design ideation using hidden Markov modeling and neuroimaging. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 1–20.

Huang, J., Jing, L., Tan, Z., & Kwong, S. (2022). Multi-Density Sketch-to-Image Translation Network. *IEEE Transactions on Multimedia*, 24, 4002–4015.

Karimi, P., Rezwana, J., Siddiqui, S., Maher, M. L., and Dehbozorgi, N. (2020). Creative sketching partner: an analysis of human-AI co-creativity. In *Proceedings of the 25th International Conference on Intelligent User Interfaces* (pp. 221-230).

Kim, J., Maher, M. L., & Siddiqui, S. (2021). Collaborative Ideation Partner: Design Ideation in Human-AI Co-creativity. *In CHIRA* (pp. 123-130).

Kwon, E., Huang, F., & Goucher-Lambert, K. (2022). Enabling Multi-Modal Search for Inspirational Design Stimuli Using Deep Learning. *Artif. Intell. Eng. Des. Anal. Manuf.*, 36(1), p. e22.

Liapis, A.; Yannakakis, G. N.; Alexopoulos, C.; and Lopes, P. (2016). Can computers foster human users' creativity? Theory and praxis of mixed-initiative co-creativity. *DCE.* Accepted: 2018-04-23T12:31:38Z Publisher: DCE.

McComb, C., Boatwright, P., & Cagan, J. (2023). Focus and Modality: Defining a roadmap to future AI-Human teaming in design. *Proceedings of the Design Society*, 3, 1905–1914. doi:https://doi.org/10.1017/pds.2023.191

Nagele, J. (2023). Fantasy on Demand: The Temptation Of Text-to-Image AI. *CTBUH Journal*, 2023(3), 46–51.

Paananen, V., Oppenlaender, J., & Visuri, A. (2024). Using text-to-image generation for architectural design ideation. *International Journal of Architectural Computing*, 22(3), 458–474.

Rezwana, J. (2023). *Towards designing engaging and ethical human-centered AI partners for human-AI co-creativity* (Doctoral dissertation, The University of North Carolina at Charlotte).

Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G. J., Elkins, A., ... Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information and Management*, 57(2), 103174. https://doi.org/10.1016/j.im.2019.103174

Stoimenova, N., & Price, R. (2020). Exploring the Nuances of Designing (with/for) Artificial Intelligence. *DesignIssues*, 36(4), 45–55.

Zhang, C., Wang, W., Pangaro, P., Martelaro, N., & Byrne, D. (2023). Generative Image AI Using Design Sketches as input: Opportunities and Challenges. 254–261.

Zhang, G., Raina, A., Cagan, J., & McComb, C. (2021). A cautionary tale about the impact of AI on human design teams. *Design Studies*, 72, 100990.

Zhou, H., Zhu, J., Mateas, M., & Wardrip-Fruin, N. (2024). The Eyes, the Hands and the Brain: What can Text-to-Image Models Offer for Game Design and Visual Creativity? *ACM International Conference Proceeding Series*.