International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hihc20

GPSdesign: Integrating Generative AI with Problem-Solution Co-Evolution Network to Support Product Conceptual Design

Pei Chen, Yexinrui Wu, Zhuoshu Li, Hongbo Zhang, Mingxu Zhou, Jiayi Yao, Weitao You & Lingyun Sun

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2025.2453003

GPSdesign: Integrating Generative AI with Problem-Solution Co-Evolution Network to Support Product Conceptual Design

Pei Chen[a], Yexinrui Wu[a], Zhuoshu Li[a], Hongbo Zhang[a], Mingxu Zhou[a], Jiayi Yao[b], Weitao You[a], and Lingyun Sun[a]

[a]College of Computer Science and Technology, Zhejiang University, Hangzhou, China; [b]School of Software Technology, Zhejiang University, Ningbo, China

ABSTRACT
In conceptual design, designers often face the challenge of navigating vast design spaces to define ambiguous problems and generate feasible solutions. Recent advancements in generative artificial intelligence (GenAI) offer new opportunities to support this process. However, formative research revealed that designers struggle to simultaneously advance both problem and solution spaces when using GenAI in conceptual design, leading to increased communication load and diminished solution practicality. This study explores the integration of GenAI with the problem-solution co-evolution model to facilitate the construction of a structured design space. We propose a GenAI-supported method for expanding and evaluating the design space and developed the GPSdesign system based on this method. Compared with a baseline system, GPSdesign fosters greater design space divergence, retrospection, and structured construction, while improving design efficiency and solution quality.

KEYWORDS
Product conceptual design; generative AI; problem-solution co-evolution

1. Introduction

Product conceptual design is the early stage of product development, where designers initially envision and articulate various aspects of a product, such as its appearance, functionality, and user experience (Benami & Jin, 2002; Li et al., 2010). This phase requires designers to apply creative thinking and design techniques to clearly define design problems and explore potential solutions in depth (Fiorineschi et al., 2016), which can be challenging. Given that this stage typically responsible for 60–80% of the costs in product development, its success is critical to the overall product outcome (Haynes & Yang, 2023).

Recent advancements in GenAI have introduced new possibilities for supporting conceptual design (Hong et al., 2023; Yüksel et al., 2023). For example, GPT-4 has a comprehensive knowledge base in the fields of design and manufacturing (Makatura et al., 2023). Designers can now utilize GenAI to obtain textual advice (Ikoma et al., 2024), as well as sketches or images (Haase et al., 2023; Shen et al., 2024). These GenAI tools help designers find inspiration (Wang, Damen, et al., 2024) and increase the diversity of ideas (Wang & Han, 2023). Furthermore, GenAI also provides design evaluation support by generating iterative reference solutions (Wang, Su, et al., 2024), analyzing user experiences (Casteleiro-Pitrez, 2024), and identifying optimal concept variants (Demirel et al., 2024).

Although GenAI is transforming the conceptual design paradigm, GenAI still faces the challenge of efficiently and accurately comprehending the intention of designers, as well as following their design process and work habit (Liu, Zhang, et al., 2024). Specifically, while the generative uncertainty in outcomes can provide diverse inspiration (Liu, Lv, et al., 2024), if not effectively managed, this uncertainty and randomness have the potential to disrupt the design process and design rhythm (Simkute et al., 2024; Yang et al., 2020). Another issue refers to the difficulty for GenAI in aligning its generated objectives with the intent of designers, resulting in inconsistent results and impeding the advancement of the design process (He et al., 2023). As tasks increase in scale or complexity, the performance of GenAI decreases, necessitating the partitioning of larger design objectives into discrete sub-goals for effective realization (Makatura et al., 2023). Therefore, integrating GenAI into the design workflow in a manner consistent with design cognitive principles can better assist designers.

The problem-solution co-evolution is a classical model for understanding the design process (Cash et al., 2023). This model characterizes the evolution of the design space as an interactive process between the problem space and the solution space, represented by four transition types: problem-problem, problem-solution, solution-problem, and solution-solution (Maher, 1990; Wiltschnig et al., 2013). During these transitions, designers continuously propose, evaluate, and reject potential design problems and solutions

CONTACT Lingyun Sun sunly@zju.edu.cn College of Computer Science and Technology, Zhejiang University, Hangzhou, 310027, China
Supplemental data for this article is available online at https://doi.org/10.1080/10447318.2025.2453003

until achieving a satisfactory "match" (Dorst & Cross, 2001). While the problem-solution co-evolution model provides a structured method for conceptual design (Dorst, 2004), designers still face significant challenges in practice. Firstly, defining design problems clearly and comprehensively remains difficult due to their vague and abstract nature, which can limit the diversity of potential solutions (Dorst & Cross, 2001; Farrell & Hooker, 2013; Fiorineschi et al., 2016). Secondly, even when following a structured process, designers invest considerable effort in identifying and utilizing valuable design representations from multi-modal and complex information (Garvey & Childs, 2015). Organizing a multi-threaded design space is essential to support effective information management and design retrospection, as co-evolution necessitates iterative reflection and logical refinement (Smits & van Turnhout, 2023).

Our research attempts to integrate GenAI with the problem-solution co-evolution model to combine the generative variability and structured design methodology. Specifically, from the perspective of the design process, the model can not only mitigate disruptions to the design flow caused by randomly generated artifacts but also provide a shared contextual environment for GenAI to enhance AI's understanding of evolution goals at each step (Tankelevitch et al., 2023). From the perspective of design information, GenAI's involvement in the co-evolution model accelerates concept generation and iteration by offering a diverse range of references (Guo et al., 2024; Makatura et al., 2023; Yu et al., 2024). The co-evolution model can efficiently manage multi-modal GeneAI artifacts by offering a multi-threaded design space (Zhou et al., 2024).

Informed by the results of a formative study with 12 design students, we proposed a generative problem-solution co-evolution method. This method integrates GenAI's expansive and evaluative capabilities with the four design transformation pathways in the problem-solution co-evolution process. Subsequently, we developed GPSdesign, a system that facilitates the generation of design problems and solutions, as well as the alignment between problems and solutions. Next, we conducted a within-subject experiment with 16 designers to verify the performance of GPSdesign. The results showed that GPSdesign can facilitate divergent thinking among designers, thereby generating more diverse and creative solutions compared to the baseline. We also visualized the design space of each designer, illustrating the differences in designers' thinking patterns under the co-evolution pattern supported by GenAI. In general, this work exhibits three primary contributions:

- Revealing the challenges designers face in the design process supported by GenAI and proposing a GenAI-supported problem-solution co-evolution design method.
- Developing GPSdesign, a system that integrates GenAI with a co-evolution model to support structured conceptual design. GPSdesign can generate multi-modal information to support creative design process, and provide decision references for the evaluation process.
- Conducting an evaluation study to verify the performance of GPSdesign in conceptual design. Moreover, we discussed the differences in design cognition and behavior among designers under the co-evolution model supported by GenAI, providing valuable insights for the study of human–AI collaboration.

2. Related work

2.1. GenAI for design

GenAI has seen rapid advancements in large language models (LLMs) and image generation models (Achiam et al., 2023). Text-based LLMs, such as GPT-4[1] and Claude 3.5 Sonnet,[2] are widely applied in fields such as writing (Law, 2024) and computer programming (Tembhekar et al., 2023), providing users with information supplementation due to their rich knowledge bases and strong-in-context understanding capabilities (Brynjolfsson et al., 2023). In the domain of image generation, diffusion models such as Stable Diffusion[3] and DALL·E 3[4] can generate high-quality images from textual descriptions, demonstrating substantial improvements in resolution, detail, and diversity.

For applications in conceptual design, the capabilities of GenAI can be classified into the following categories: inspiration generation, idea and concept generation, and concept evaluation (Morris et al., 2023; Yüksel et al., 2023). For inspiration generation, GenAI typically generates multi-modal referential content based on the specific requirements of the design task (Jiang, 2024). For example, InspireMe generates textual descriptions, images, and virtual models for designers (Zhang, Cai, et al., 2024). The simultaneous presentation of multiple representations has been shown to foster deeper insights and facilitate broader exploration compared to focusing on a single representation (Yao et al., 2024). For idea and concept generation, GenAI fosters a co-creative relationship with designers, necessitating shared context for effective collaboration. For example, CausalMapper (Huang et al., 2023) broadens design exploration by generating potential problems or solutions based on existing design networks, which designers refine using their expertise. Zhong et al. (2024) introduced causal path diagrams and used LLM to generate preconditions, thus facilitating divergent thinking. ProtoDreamer (Zhang, Chen, et al., 2024) combines GenAI with physical prototyping, allowing designers to construct preliminary prototypes while AI generates diverse design alternatives. For concept evaluation, GenAI assists in assessing and improving design content based on the contextual information. For example, MemoVis helps create reference images of 3D designs that align with text feedback, improving evaluation efficiency and clarity (Chen et al., 2024).

However, current GenAI-supported design assistance tools still have limitations when handling complex tasks. First, output uncertainty can disrupt designers' original ideas and logical flow. While these tools can encourage exploration (Morse et al., 2018), the inconsistent outcomes it generate can lead to unworkable design concepts or serious production disruptions (Alam et al., 2024). Second, it is difficult to align GenAI's design objectives with those of designers (Saadi & Yang, 2023; Sun et al., 2024). Due to the inherent abstraction and complexity of design activities, designers often find it challenging to convey their requirements and constraints precisely through brief descriptions, as these elements are frequently embedded within the design context and multi-modal materials (Chung & Adar, 2023). Additionally, the design process is continually evolving, with designers discovering new problems and generating new ideas. Current design tools struggle to enable AI to capture the critical information that designers focus on within a nonlinear, unstructured process and to follow the designer's thought process to provide targeted support (Alam et al., 2024). Therefore, it is necessary to develop AI-assisted design tools that align with the inherent principles of design activities.

2.2. Problem-solution co-evolution in conceptual design

The problem-solution co-evolution model defines the design process as a dynamic iterative process where problems and solutions co-constrain and evolve together (Maher & Tang, 2003). Designers continuously propose, evaluate, and reject potential design problems and solutions until they can create a good "match" between the problem space and the solution space (Dorst & Cross, 2001). The "matching" process is evaluated by a fitness function, which is designed to identify pairs of problem-solution nodes that exhibit high fitness as being better matched (Maher et al., 1996). Taking steel frame design as an example, the problem space may include representations such as structural efficiency, structural integrity, and compatibility with architectural layouts. The solution space may include representations such as the geometric dimensions of the frame and the cross-sectional dimensions of components. The fitness function can be designed to assess structural efficiency, structural integrity, and compatibility with architectural layouts.

Previous research has highlighted the applicability of this model in many design fields (Crilly, 2021a, 2021b), such as industrial design (Crilly & Firth, 2019), innovative product design (Hui et al., 2020), parametric design (Yu et al., 2015), etc. For example, in the field of industrial design, Dorst (2019) presented a real-world project focused on designing child safety seats to illustrate the iterative and high-level co-evolutionary characteristics of this model. Co-evolution is also a crucial collaborative model in the design process, playing an important role not only in collaboration between designers and clients (Smulders et al., 2009) but also within design teams (Cash et al., 2023; Wiltschnig et al., 2013). However, limited studies have explored the co-evolution model and interaction types between designers and GenAI (Guo et al., 2024).

The co-evolution model may offer the following advantages to the human–GenAI co-design. First, the co-evolution model can mitigate disruptions in design flow caused by randomly generated artifacts. For example, Fiorineschi et al. (2016) integrated the co-evolution model to propose a Problem-Solving Network (PSN). This network aims to enhance the management of information-gathering activities, which is essential for effective problem-solving processes. Each branch within the PSN is independent, which means that the evolutionary trajectory of each problem-solution pair can develop autonomously. Therefore, we can assume that when GenAI in the branch breaks or produces errors, the negative impact will not interrupt all design workflows. Second, the co-evolution model may compensate for GenAI's limitations in understanding and identifying design problems. Martinec et al. (2020) noted that co-evolution shows a significantly higher effect on synthesizing new problem entities than on generating new solution entities. This advantage is particularly beneficial for incorporating GenAI into the design process, as it enables designers to convey design problems more clearly. Previous research has shown that the more detailed and focused the description of the design problem, the better GenAI performs (Makatura et al., 2023).

Despite these potential advantages, how designers use GenAI to identify problems and determine solutions, as well as what challenges they encounter, remains unclear. In this study, we first conducted a formative study to identify the challenges designers face when using existing generative tools to complete design tasks. Based on the results, we proposed a method that integrates GenAI with a problem-solution co-evolution network, leading to the development of a design support system.

3. Formative study

We conducted a formative study to investigate the challenges faced by designers in using GenAI to identify problems and determine solutions during the conceptual design process. The findings of this study provided a reference for the design of our system.

3.1. Method

3.1.1. Participants and procedure

We conducted an experiment involving 12 participants ($M = 22.42$, $SD = 1.19$; five males, seven females) for this formative study. All participants had at least two years of experience in conceptual design and were familiar with GenAI tools such as ChatGPT,[5] Midjourney,[6] and Stable Diffusion WebUI.[7] The experiment employed a web-based canvas that supported text editing and doodling for conceptual design tasks. Participants were able to seek AI assistance through the official OpenAI ChatGPT website platform, which utilized the GPT-4 model for text conversations and DALL·E 3 for text-to-image generation. Details of the interface and other aspects of the experimental system can be found in Supplementary Appendix A.1.

We prepared two open-ended tasks with different themes: designing a mobility aid for the elderly and designing a smart home cleaning product. These tasks were chosen to eliminate experimental bias caused by task specificity and to facilitate designers' use of prior experience. Additionally, the tasks required designers to consult additional resources for specific functional and structural design methods, encouraging them to seek assistance from AI rather than relying solely on their knowledge. This approach allows us to observe and record designers' processes when using AI assistance.

We randomly divided participants into two groups. One group completed task one, and the other group completed task two. Before the experiment, participants signed an informed consent form. After introducing the experimental process and requirements, participants completed the design tasks according to instructions. The entire experiment lasted about 50 min. In each design task, participants had 24 min for idea expansion and 8 min to finalize their concepts. After completing the design tasks, we interviewed them about their typical conceptual design practices, challenges in identifying design problems and solutions, and their experiences with the AI tools. The interview process was audio-recorded. We also recorded the conversation logs between the participants and AI, which contained the participants' prompts and the images and text generated by GPT-4 and DALL·E 3.

3.1.2. Evaluation metrics and data analysis

To assess the quality of participants' design task outcomes, we invited three design experts to evaluate the creativity of their design proposals. Each expert possesses at least six years of experience in conceptual design and is capable of making independent and professional judgments on design solutions. We referred to metrics based on prior research (Hwang & Park, 2018; Sarkar & Chakrabarti, 2011), which assess creativity by evaluating originality and practicality. Specifically, practicability (P) is calculated as the level of importance (L), the rate of popularity of usage (R), and the frequency of usage (F):

$$P = L \times R \times F$$

More details can be found in Table 1. After obtaining the originality and practicality scores, we performed a normalization process to facilitate further analysis.

For the interview analysis, we transcribed the audio recordings and extracted participants' comments on the AI-assisted design. Two coders (co-authors of this article) analyzed these comments using a hybrid coding approach (Winske & Omidi, 1991). Focusing on the advantages and disadvantages of using current AI tools in the design process, the coders independently assigned initial codes to the comments and then discussed and refined these codes to resolve any discrepancies. The final codes included categories such as "providing solution details," "expanding ideas," "promoting reflection," "not integrated into the design process," "information repeated," and "information deviating from the topic." Relevant excerpts are cited in Section 3.2.

For the conversation logs between participants and AI tools (GPT-4 and DALL·E 3), we focused on how participants drew inspiration from AI and why AI sometimes failed to generate expected information (Supplementary Appendix A.2 presents an example).

For the design outcomes, we examined all design proposals submitted by the participants. Some proposals clearly articulated the design problems, while others focused solely on functional aspects. Specific examples are provided in Supplementary Appendix A.3. We considered a participant to demonstrate co-evolutionary design thinking if more than half of their proposals contained well-defined design problems. Detailed analysis findings are presented in Section 3.2.2.

3.2. Findings

Our results highlight the crucial role of clear problem definition in effective communication between AI and participants. Vague problems often lead to inaccurate or ambiguous information generation. Balancing attention between problems and solutions enhances the practicality of design outputs. Additionally, design information presented in text and images modalities each offer distinct advantages. We summarize our key findings as follows:

3.2.1. Problem definition affects the communication between GenAI and designers

Participants struggled to obtain detailed information from AI when they did not clearly define design problems. Most participants frequently recorded all feasible design solutions on the canvas, paying limited attention to the design problem itself. Interviews revealed that this oversight does not originate from a lack of consideration but rather from a predisposition to propose specific solutions, holding the key problem statements internally. For example, P10 stated: "*I will not describe too many requirements in the proposal statement, but I believe that my different solutions address different needs.*" This internalization led participants to rely on their memory when discussing with GPT-4 or requesting image generation from DALL·E 3, making it difficult for AI to fully and accurately understand the current design task's challenges. Consequently, the participants' general and vague queries hindered AI's ability to provide responses precisely aligned with their immediate design goals.

Table 1. Expert evaluation criteria for design proposals in the formative study.

Dimension	Sub-dimension	Statement
Originality	–	Whether the idea involves innovative and original features compared to existing products, assessed on a seven-point scale (1 = not novel at all, 7 = extremely novel).
Practicality	Level of importance	Whether the idea involves an indispensable nature in enhancing the user experience or solving problems of target scenarios, assessed on a five-point scale (1 = not important at all, 5 = extremely important).
	Rate of popularity of usage	The proportion of target users likely to use the product, assessed on a scale from 0 to 1, rounded to one decimal place (0 = no one would use it, 1 = everyone would use it).
	Frequency of usage	The estimated frequency of target users' product usage, assessed on a scale from 0 to 1, rounded to one decimal place (0 = never used, 1 = used every time).

3.2.2. Balanced consideration of problems and solutions enhances the practicality of design outputs

The analysis of design outcomes indicates that only six participants consistently identified the targeted design problems. They considered the correspondence between the design problem and the solution.

To further analyze the impact of balanced consideration of problems and solutions on design outcomes, we calculated the relationship between expert scores for solutions generated by participants with and without co-evolutionary design thinking. As shown in Figure 1, participants with co-evolutionary design thinking achieved a significantly higher average practicality score ($p = 0.010$), while no significant difference was observed in originality between the groups ($p = 0.813$). This insignificance likely stems from GPT-4's extensive knowledge base, and it infers unspecified aspects in prompts (Makatura et al., 2023). Thus, it can associate keywords with a broad array of related scenarios, applications, and functions, providing participants with diverse creative ideas. However, when problem constraints are vague, AI responses often remain generalized, consistent with views of Rasal and Hauer (2024). Consequently, expert evaluations showed that while solutions from all participants included novel concepts, those generated with co-evolutionary design thinking were more specific and practical. Participants who balanced problem and solution considerations were also more selective, choosing the most feasible schemes from AI-generated suggestions. As P8 noted, "Although the AI provides many options, it is unrealistic to implement all of them within a single product. I would select some ideas to combine into the most practical solution." Thus, in AI-assisted design, balanced consideration of problems and solutions allows the diverse concepts provided by AI to be developed into feasible design solutions rather than remaining at a superficial level.

3.2.3. Participants utilize AI to generate different modalities of information at various stages of the design process

Text and image modes have certain differences in the accessibility of information, affecting their usage and reference value throughout the design stages. Consistent with previous research, our interviews indicated that the text mode is more intuitive and specific for describing the requirements and functionalities of the problem, while the image mode can quickly convey information about the appearance, structure, and materials (Zhang, Cai, et al., 2024). Throughout the design process, the generation of textual information was continuous. However, as designs progressed, the requests for image generation from participants increased. Participants explained that in the early stages of design, the product concept was vague, and images generated by AI based on these vague concepts seemed reasonable but provided far less information about problem insights and functionalities compared to text. In later stages, with clearer design concepts, AI can generate specific product images from design descriptions, aiding participants in evaluating appearance and structure feasibility and fostering iterative development. Moreover, although AI can create illusions, these illusions were sometimes viewed as inspirational when used merely as design references (Lee, 2023; Mbalaka, 2023). For example, when a participant requested to generate an image of a smart window-cleaning device, the returned image resembled a drone, inspiring the participant to develop a high-rise window-cleaning machine concept based on drone technology.

3.3. Design goals

Based on the study findings, providing a detailed description of the design problem enables AI to comprehensively understand design tasks. Balancing consideration of the relationship between design problems and solutions enhance design practicality. We identified three design objectives to establish a system that integrates AI capabilities with problem-solution co-evolution model to support the conceptual design process:

- DG1: Introducing the problem solution co-evolution model to explicitly promote designers' balanced consideration of the problem space and the solution space.
- DG2: Enabling AI to follow the design process and to assist designers in finding optimal solutions and problems within a complex design space.
- DG3: Using multi-modal generative content to support the evolution of design space, while reducing the communication burden between designers and AI.

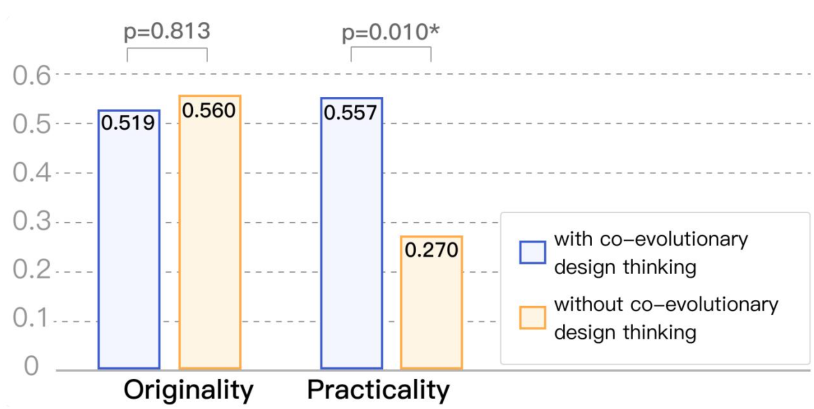

Figure 1. Expert rating results on originality and practicality in the formative study.

4. Generative problem-solution co-evolution method

Aligned with the design objectives, we propose a generative problem-solution co-evolution method that integrates AI capabilities with the problem-solution co-evolution model to comprehensively support the conceptual design process. Our method portrays the design space as a network and includes the strategy for facilitating the expansion of the design space (see Figure 2) and the strategy for supporting the evaluation of design space (see Figure 3). In previous co-evolution models, node transformations include problem evolution (P-P), proposing solution (P-S), proposing problem (S-P), and solution evolution (S-S) (Gero et al., 2022). In practical design activities, these transformations do not occur strictly linearly but happen concurrently. In our method, we further divide problem and solution nodes into designer problem nodes, designer solution nodes, AI problem nodes, and AI solution nodes. This categorization further clarifies how AI participates in design activities. In addition, by understanding the problem-solution network, AI can follow the design process and support designers by adopting different strategies in the design expansion and evaluation stages.

4.1. Strategy for facilitating the expansion of the design space

AI can rapidly expand problem and solution spaces when designers seek inspiration. Generated references include:

- A—Generating solutions from problems (P-S)
- B—Generating new problems from problems (P-P)
- C—Generating new solutions from solutions (S-S)
- D—Generating problems from solutions (S-P)

Activity A and D represent the occurrence of co-evolutionary events where the designer alternates between focusing on the problem space and the solution space (e.g., discovering design problems to be solved from the solution and proposing new solutions to solve the problems). Activity B occurs when designers focus on the evolution of the problem space (e.g., refining the problem). Activity C occurs when designers focus on the evolution of the solution space (e.g., functional decomposition). For instance, a designer may initially propose a concept, and AI can generate a series of new questions based on this concept, such as "how to improve the product's sustainability" and "how to reduce manufacturing costs?" assisting designers in approaching problems from various perspectives. Simultaneously, AI can also derive new solution concepts from this initial idea, such as incorporating novel materials or processes, to inspire designers. Furthermore, issues like "cost" and "sustainability" may evolve alongside the updates in solutions related to "materials" and "technologies." Consequently, specific solutions for new design problems are subject to change.

By applying this classification, we systematically deconstruct the specific behaviors that AI can augment. After identifying these behaviors, we need to determine the optimal number of results generated in each iteration. Existing studies indicate that presenting multiple different design solutions might lead to better design outcomes and bolsters creativity and self-efficacy (Dow et al., 2010; Wadinambiarachchi et al., 2024). However, presenting too much information might increase the cognitive load. Therefore, we specified to generate three distinct results each time, balancing diversity with information management.

4.2. Strategy for supporting the evaluation of design space

Previous research has established that the process of creative design encompasses an exploratory phase, during which the

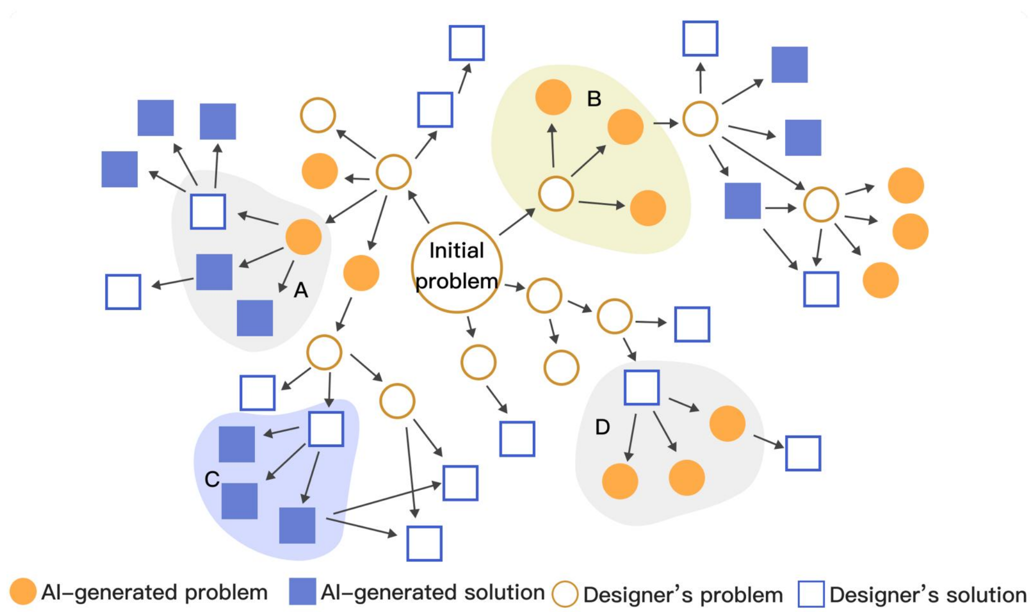

Figure 2. The strategies for facilitating the expansion of the design space. (A) Generating solutions from problems; (B) Generating new problems from problems; (C) Generating new solutions from solutions; (D) Generating problems from solutions.

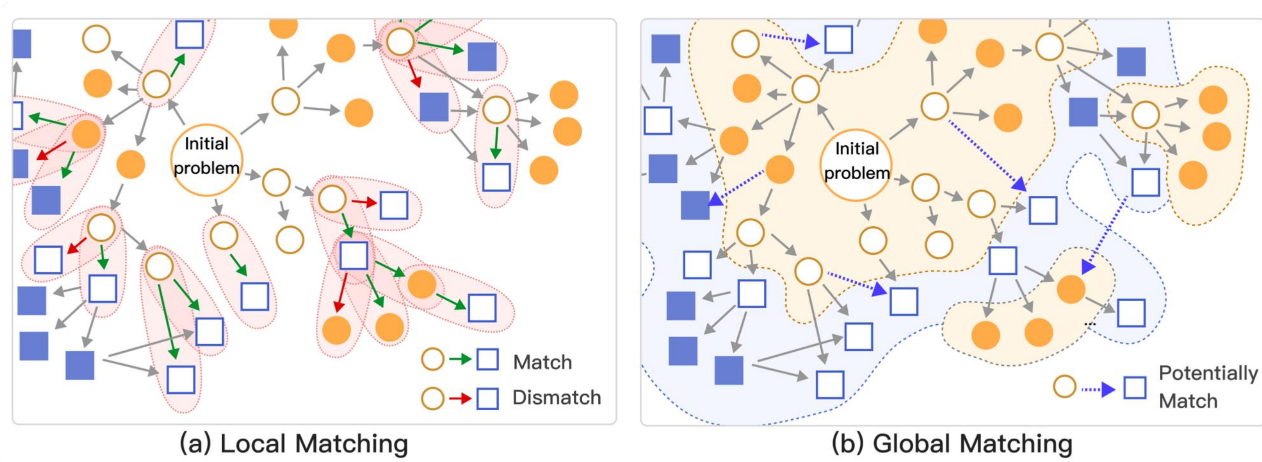

Figure 3. Strategies for evaluating problem-solution matches. (a) The local matching method assesses the alignment between connected problem-solution pairs; (b) The global matching method assesses all problem and solution nodes in the design space, identifying potential matches between problems and solutions that the designer has not found.

problem and solution space undergo continuous evolution and instability until a congruence between the problem and solution is established (Dorst & Cross, 2001; Schön, 1992). Therefore, identifying those matching problems and solutions within the design space is a crucial step toward achieving creative design outcomes. However, in the later stages of the design process, the design space often becomes highly complex, making it challenging for designers to conduct rapid reviews and iterations.

We propose employing AI to assist designers in evaluating design spaces at both local and global levels. In the fundamental co-evolution model, researchers have compared the evolutionary approaches of solution and problem spaces to biological mechanisms: the *Combined Gene Approach* and the *Interacting Population Approach* (Maher et al., 1996; Maher & Wu, 1999).

- *Combined Gene Approach*: Problems and solutions are integrated into a single composite genetic type, with their fitness defined locally for each design solution. Node pairs with low matching degrees are eliminated. This approach resembles tightly coupled or host-parasite co-evolution.
- *Interacting Population Approach*: Fitness evaluation of two spaces is conducted alternately, representing loosely coupled or predator-prey co-evolution. The best individuals serve as benchmarks for the next generation of individuals. For example, the solution space adapts to the most important problems in the existing problem space.

Inspired by these two approaches, we proposed a local matching method (see Figure 3(a)) and a global matching method (see Figure 3(b)).

- **Local matching:** Employing AI to assess whether the connected problem-solution pairs match. This means checking if each connected problem-solution node pair is suitable while ignoring any unconnected pairs.
- **Global matching:** Employing AI to assess whether there are potential matches between problem-solution pairs within the design space. In other words, identify problem and solution node pairs in the design space that are not explicitly connected by the designer but still exhibit a match. In the later stages of conceptual design, the solution space needs adjustment to address critical issues in the problem space. Due to the inherent vagueness and uncertainty of the design process, pinpointing the most important problems and solutions is often difficult (Daalhuizen et al., 2009). Factors that seem insignificant can suddenly become crucial. Identifying these potential matches helps designers uncover overlooked elements or explore pivotal design paths.

5. GPSdesign

5.1. System design

Building on the proposed generative problem-solution co-evolution method, we introduce GPSdesign, a GenAI-supported tool that utilizes node-based associative maps to structurally assist designers in advancing conceptual design. It can generate multi-modal information to support creative design processes and provide decision references for the evaluation process. As shown in Figure 4, GPSdesign is presented as a web tool. The main canvas supports the addition of various types of nodes, with specific node types controlled in the edit area. GPSdesign supports two stages of the design process: ideation expansion (see Figure 4(1)) and evaluation convergence (see Figure 4(2)). We detail the specific functionalities provided at each stage as follows:

5.1.1. Ideation expansion phase

In the ideation expansion phase, designers can structurally broaden both the problem and solution spaces. During this phase, designers have the option to add "AI-assisted Design" node (see Figure 4(c)) and "My Design" node (see Figure 4(d)) to the canvas. The initial description of a design task

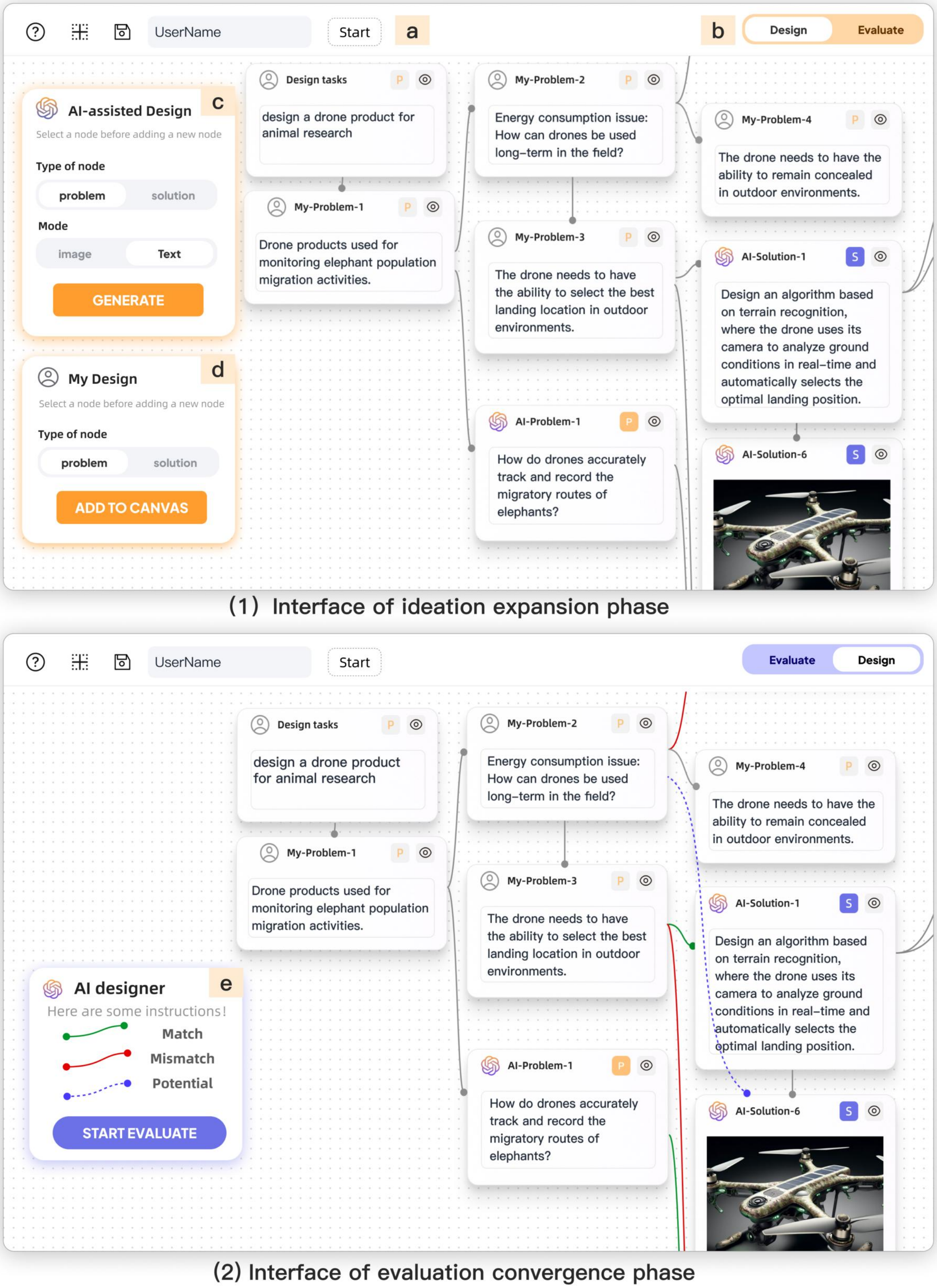

Figure 4. The main interface of GPSdesign. (1) Interface of ideation expansion phase; (2) interface of evaluation convergence phase. (a) Operation guide and user information input; (b) button to switch user interface; (c) edit component for adding AI nodes; (d) edit component for adding custom nodes; (e) instructions and trigger buttons for evaluation.

serves as the starting node for the entire node map. When designers intend to add new node to the canvas, they need to designate a node as the parent node. The newly added node will automatically link to its parent node, maintaining a clear structural flow. For adding their creative ideas to the canvas, designers can use the "My Design" edit area to construct text-based problem or solution nodes. If designers encounter a lack of inspiration, they can actively request AI-generated problem or solution nodes. The function for adding AI nodes allows designers to specify node types and content modalities, providing three new nodes for each request. In summary, GPSdesign can generate four types of nodes:

1. AI problem nodes in text modality.
2. AI solution nodes in text modality.
3. AI problem nodes in image modality, which generate scene images that depict the problem.
4. AI solution nodes in image modality, which generate accurate product images that depict the solution.

5.1.2. Evaluation convergence phase

When designers are ready to enter the evaluation phase, they can simply click on the toggle button in the upper right corner to access the evaluation interface (see Figure 4(2)). In the evaluation phase, designers click on the "Start Evaluation" button (see Figure 4(e)) and GPSdesign will perform two types of evaluation tasks:

1. Assessing the compatibility of the connected problem and solution node pairs; compatible connections are indicated in green, while incompatible ones are marked in red.
2. Evaluating potential matches between solutions and problems. If a pair of problem-solution nodes that are compatible but not connected is identified, a connection will be established between the two.

After the evaluation, designers can utilize the results for reflection on the design space and ensuring precise definition of design concepts. If designers are not satisfied with the current design space, they can go back to the ideation expansion phase to improve their design.

5.2. System implementation

We implemented GPSdesign using Python Flask[8] and ReactJS.[9] We adopted MongoDB[10] to store designers' personal information and the multi-modal information (text and images) generated during the design process. Specifically, we called GPT-4 turbo[11] API for text generation and the DALL·E 3[12] API for image generation. Based on guidelines established for large language model prompts (Zhao et al., 2023) and all possible node transformation actions within the system, we implemented 16 distinct generation prompt strategies, each tailored to a specific node transformation characteristic. Additionally, we developed a method to evaluate the compatibility between all problem nodes and solution nodes within the design space. The following sections will discuss these details and the complete list of prompts is shown in Supplementary Appendix B.

5.2.1. Method for node generation

Section 5.1.1 has delineated the four types of AI nodes. The process of generating nodes involves creating child nodes from a parent node, necessitating consideration of a total of 16 node generation strategies in the design process when considering both parent and child node types. These 16 strategies can be categorized under the AI extension capabilities categories mentioned in section 4.1.

The prompt engineering for content generation includes a basic template. Figure 5(a) shows a case of generating a text solution node from a text node. The template comprises the task introduction, task requirements, one-shot examples, output specifications, and input content.

If an image node is required to be generated, the initial step involves creating new text content, such as product concepts for solution nodes or scene concepts for problem nodes. Subsequently, text prompts are generated based on the specified image prompt requirements. Finally, DALL·E 3 generates images in accordance with the prompts and provides both the image results and newly generated text content to the front-end. Figure 5(b) shows a case of generating an image solution node from a text problem node. Prompt engineering varies depending on different nodes and modal characteristics, thus detailed content will be elaborated in subsequent sections.

5.2.1.1. Generate problem nodes. If problem nodes in the text modality need to be generated, one requirement is that each node should contain no more than 50 characters to reduce the cognitive load on users to understand each piece of text. Additionally, it is important for the content between nodes to be as distinct as possible to enhance divergent thinking. The example we provide requires that GPT-4's description be presented in the form of a question, with a primary focus on abstract dimensions, such as questioning the product's safety, durability, and other dimensions. However, in practice, when the content of the parent node is more specific, the questions generated by GPT-4 will also become more focused and detailed.

If problem nodes in the image modality are required, the first step is to ask GPT-4 to generate three different related scenes, where the scenes need to be in specific locations and have significant differences between them. Additionally, the content of the image prompts should include descriptions of the details in the scene, scene lighting, and perspective, with as many details as possible.

5.2.1.2. Generate solution nodes. If solution nodes in text modality nodes are required to be generated, the node's word count should not exceed 80 words. Similarly, it is important that the content between nodes differ as much as possible.

If solution nodes in the image modality are required, the first step is to request AI to produce three distinct product concepts (not functional descriptions) to ensure that the generated images are relevant to a specific product form. Simultaneously, it is essential for the content of the image prompts to encompass descriptions of the item's appearance, material, shape, structure, and lighting with as many details as possible.

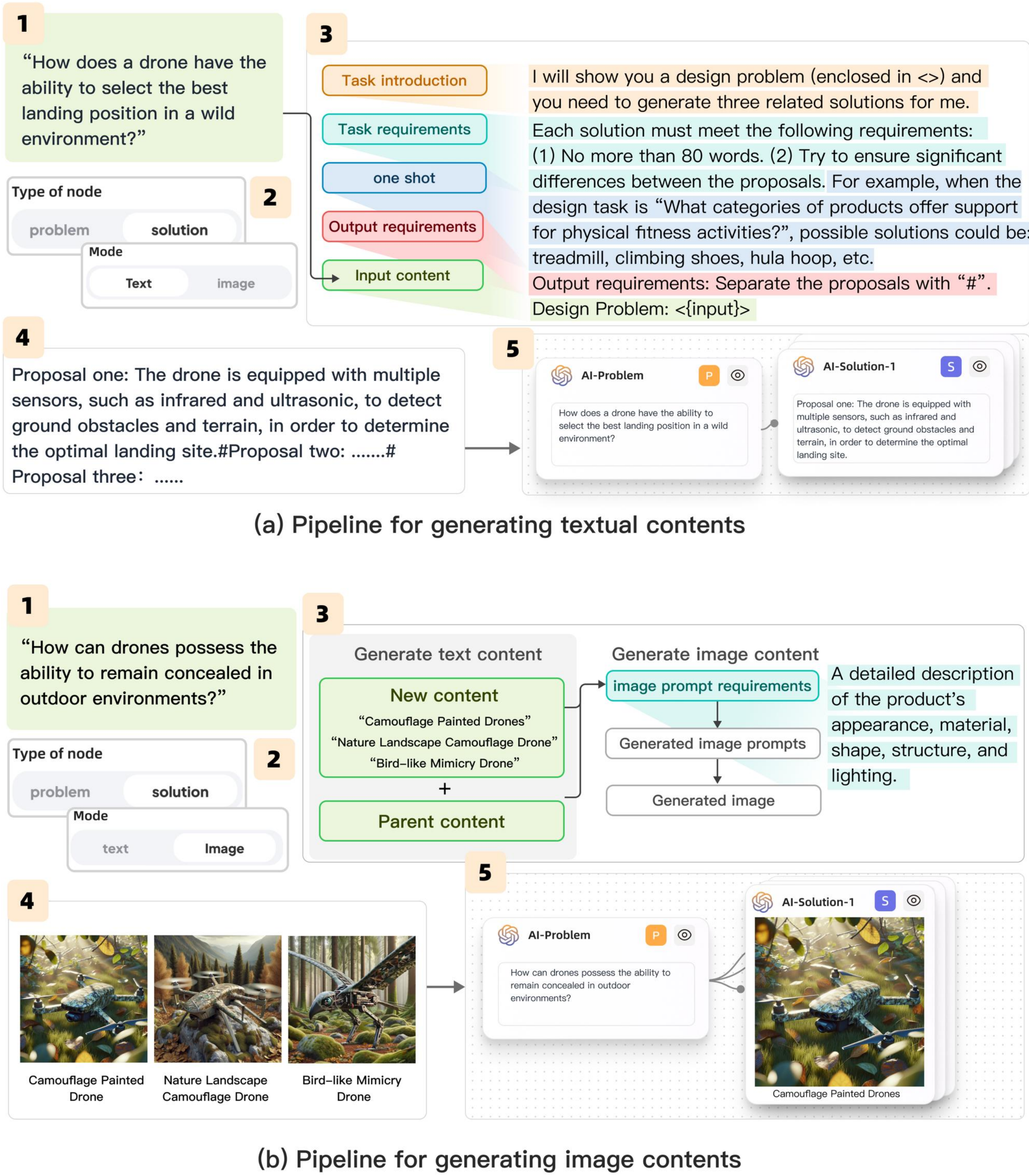

Figure 5. A case of generating solution nodes from problem nodes. (a) Pipeline for generating textual content; (b) pipeline for generating image content.

5.2.2. Method for evaluation

During the construction of the design network, the front-end data records all the content of the problem nodes and solution nodes, as well as each connected problem-solution node pair. These content are stored in JSON format. For example, a problem node is recorded with a unique identifier ("ProblemID") and specific problem content ("ProblemData"), such as "My-Problem-1" and "How is the durability of drones and their ability to adapt to various environments?"

The evaluation pipeline is illustrated in Figure 6. We first provide GPT-4 of all content of the problem nodes and solution nodes without revealing the connections between these nodes. GPT-4 then identifies matching problem-solution pairs based on two criteria: (1) Does the solution solve the design problem? (2) Is the problem related to the design solution? Meeting either criteria suggests compatibility and GPT-4 returns matches with a brief justification. These are recorded in JSON format with fields for "ProblemID," "SolutionID," "MatchStatus," and "Reason." Following this, the front end processes the evaluations according to the following logic:

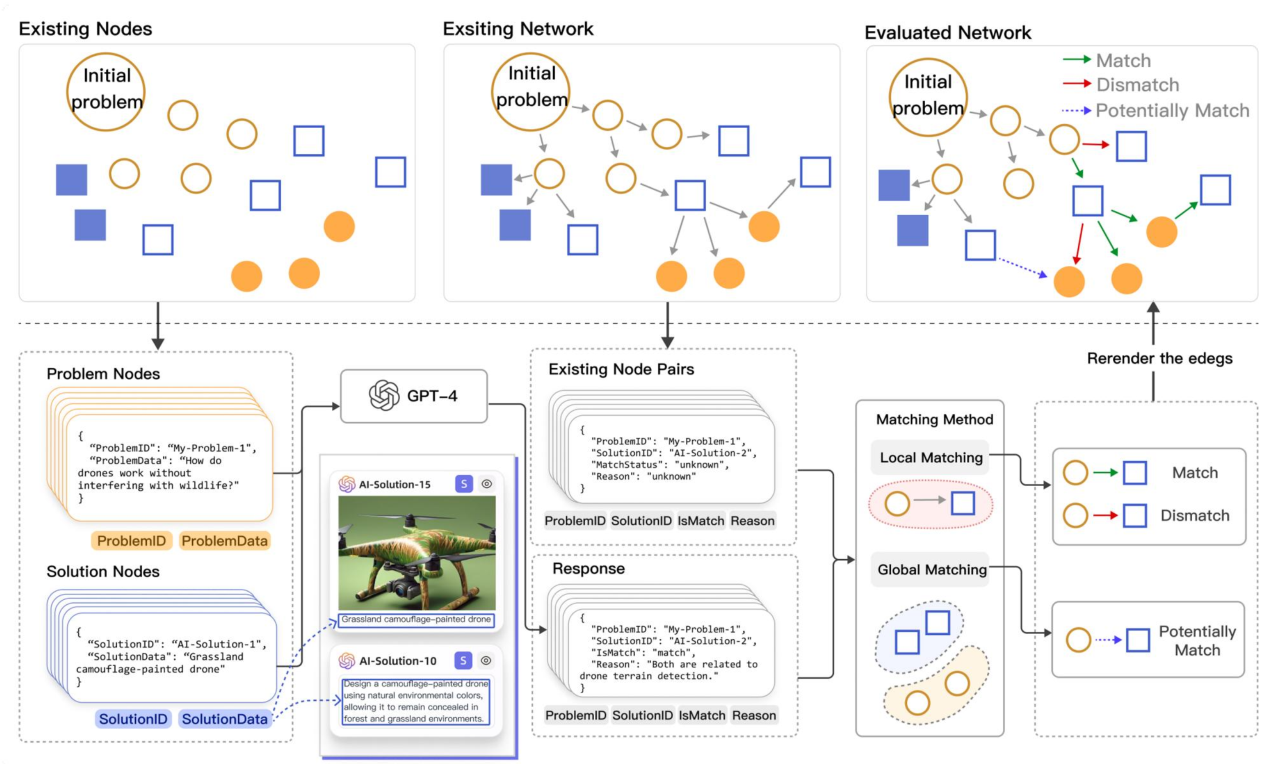

Figure 6. Evaluation pipeline of GPSdesign.

5.2.2.1. Local matching. Existing connected node pairs are compared to GPT-4's results. Connected pairs judged as incompatible are marked in red, while compatible pairs are marked in green.

5.2.2.2. Global matching. Compatible but unconnected pairs identified by GPT-4 are highlighted with a blue dashed line, indicating potential matches for exploration.

6. User study

To examine the impact of GPSdesign on the design process and outcomes, a within-subject experiment was conducted with 16 participants. The study aimed to address the following research questions.

- RQ1: Is GPSdesign usable? Can GPSdesign provide the desired support for designers during their design process?
- RQ2: Does GPSdesign enhance creative ideation during the design process?
- RQ3: How does the generative co-evolutionary mode of GPSdesign influence participants' design thinking patterns?

6.1. Participants

We recruited 16 participants (eight females, eight males; M = 23.5, SD = 1.32). Among the participants, 14 were studying industrial design and had at least three years of design experience, while the remaining two were mechanical engineering students with over five years of experience in mechanical engineering. All participants were proficient in utilizing GPT-4 to generate images and text and had prior experience with web-based mind-mapping tools.

6.2. Procedure

In the experimental process, participants were required to complete two concept design tasks using GPSdesign and the baseline system, separately. Task A was to design a robotic arm product for use in the agricultural field. Task B was to design a drone product for animal research. The two tasks were designed with similar descriptions and requirements to minimize task-related variations. The baseline system offered node addition and editing functions but lacked node categorization and matching evaluation capabilities. To control for variations in model generation effects, participants in the baseline condition also used the same generation models (GPT-4 for text and DALL·E 3 for images) for assistance. More details about this baseline system are provided in Supplementary Appendix C.1.

To mitigate potential biases related to task sequence or condition effects, we counterbalanced the order of conditions and tasks among the 16 participants. Figure 7 illustrates the experimental procedure. Participants were randomly divided into four groups and completed two sessions of design tasks. Each section consists of a task briefing, tool instruction, ideation phase, post-task survey, and interview. During the ideation phase, participants used the assigned system to complete the thematic design tasks within 20–30 min. After the ideation phase, participants

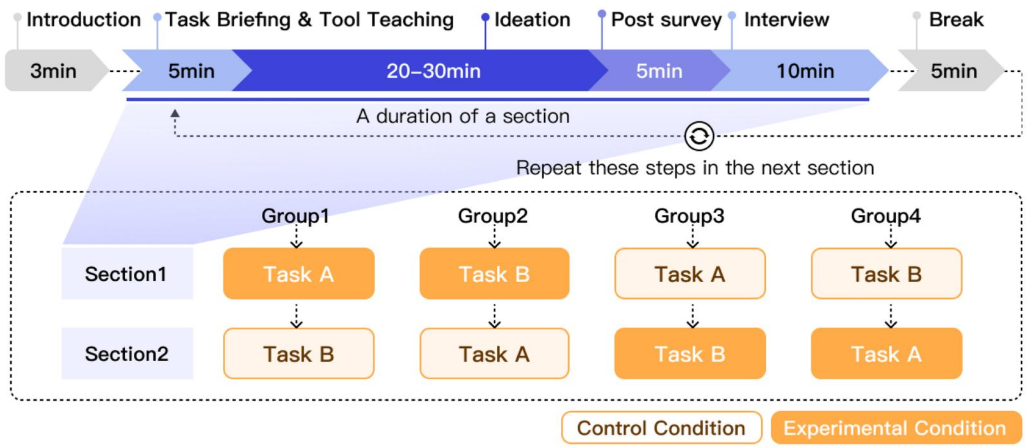

Figure 7. Experimental procedure of the user study.

Table 2. Self-customized questionnaire.

Dimension	Question
Expectation	The content generated by the tool meets my expectations.
	I can integrate the content generated by the tool into my design well.
Modality	Multi-modal generated content helps me explore ideas.
	Multi-modal content is more helpful than single-text content.
Efficiency	This tool has increased my efficiency.
	Using this tool, I can produce more and higher quality ideas.
Insight	This tool helps me think deeply about problems.
	This tool helps me iterate solutions.

completed a post-task survey (details in Section 6.3) and participated in a 10-min semi-structured interview to explore experiential differences between the two sessions. A 5-min break was provided between the two sections. The study lasted 90–110 min, and each participant received 90 RMB as compensation. All participants signed a consent form approved by our institution.

6.3. Measures

For usability assessment, we used the system usability scale (SUS) questionnaire (Brooke, 1996). It evaluates the overall usability of a system in terms of learnability and effectiveness, using a 5-point scale. To assess the system's support for creativity, we adopted the Creativity Support Index (CSI) (Carroll et al., 2009; Cherry & Latulipe, 2014). The CSI measures six dimensions of creativity: exploration, expression, immersion, enjoyment, results worth effort, and collaboration, using a 10-point scale. Especially, since our system and the baseline do not involve multi-user collaboration, we excluded the collaboration dimension.

To investigate other specific aspects of system support mentioned in RQ1, we developed a custom-designed questionnaire (see Table 2) based on findings from the formative study. Specifically, we created eight questions to assess the effectiveness of our system across dimensions such as "meeting expectations," "usefulness of multi-modal information," "enhancing efficiency," and "providing insight." The questionnaire utilized a 7-point scale.

To understand the design thinking patterns supported by GPSdesign, we recorded the design process of the participants to understand their creative thinking and interactions with AI. To examine differences in exploring design spaces under different conditions, we encoded the frequency and characteristics of node transformation patterns within each participant's design process, segmenting the design phases over time. The specific coding methods are as follows:

- **Design process coding method:** The method captures the transformational actions that occur during the design process, which encompasses four distinct types: adding a solution based on a problem (P-S), adding a problem based on a problem (P-P), adding a problem based on a solution (S-P), and adding a solution based on a solution (S-S). In particular, since there is no distinction between the problem and solution nodes in the baseline, we established standards for distinguishing them. A design problem is considered any "problem" expressed as "how to verb-noun?" or in a "gerund" form, which might not necessarily indicate a function but could refer to what the system should do according to specified design requirements (Fiorineschi et al., 2016). A design solution is viewed as an answer or a pathway to solve the posed problem, detailing how to address and resolve the issue. It usually includes the solution strategy, implementation steps, and required resources. The coding schemes and results are detailed in Supplementary Appendix C.3.
- **Design results coding method:** To capture the characteristics of participants' design spaces, we visualized the design process manually using radar-chart-like representations inspired by the work of Kim and Kim (2015). We encoded nodes in the design spaces according to the scheme in Figure 8. In our scheme, each circular layer represents an iterative evolution of problems or solutions, and the clockwise direction represents the order of node addition. We highlighted key nodes that participants considered highly relevant and inspiring, determined by them during the experiment. Gray dotted lines show backtracking paths of key nodes, clarifying their formation process. Note that the innermost layer includes iterations of the initial problem space; addition of solution nodes starts from the second layer. This format clarifies

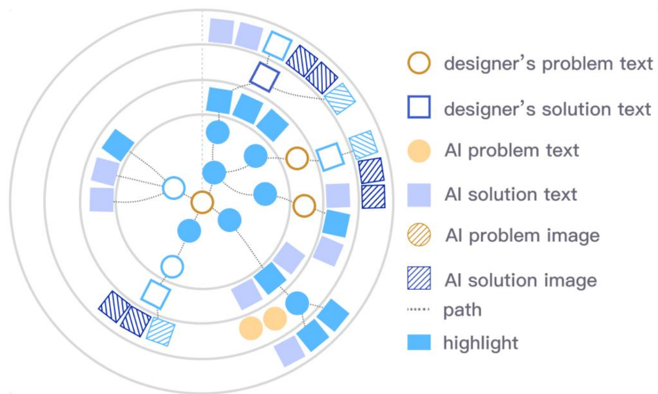

Figure 8. The encoding scheme for the design space.

the separation between the initial problem space and the initial solution space.

7. Results

The overall results indicate that AI-generated content under the co-evolution mode can encourage participants' divergent thinking, promoting designers to uncover a wider variety of problems and solutions, thereby helping designers to establish a richer design space. We also found that there are significant differences in the creative processes of different participants when designing with GPSdesign. Examples of the design process and outputs with GPSdesign and the baseline system are shown in Supplementary Appendix C.2.

7.1. RQ1: GPSdesign is usable and provides the desired support for designers

To answer RQ1, we examined the results of the SUS and the self-customized questionnaire. The content of the interviews served as supplementary explanations.

The overall average SUS scores for GPSdesign and baseline were 80.2 ($SD = 9.42$) and 72.7 ($SD = 12.47$), respectively. Both scores exceed the SUS acceptance threshold of 70, denoting satisfactory user acceptance. Notably, GPSdesign's overall usability significantly outperformed the baseline ($p = 0.005^{**}$), though it exhibited a significantly lower learnability score compared to the baseline ($p < 0.001^{***}$). Participants are more familiar with the baseline system. This is due to prior exposure to similar systems such as GPT-4, which aligns more closely with their existing experience. However, feedback from participants suggested that the learning curve for GPSdesign remained within an acceptable margin. P14 remarked, *"I feel that the baseline is easier to get started with, but the other one is not difficult to operate either; it just has more functions, and I can also get the hang of it quickly."*

Results from our self-customized questionnaire indicate that GPSdesign significantly outperforms the baseline across the dimensions of modality, efficiency, and insight (see Figure 9). The modality scores of GPSdesign ($M = 6.16, SD = 0.85$) are significantly higher than that of the baseline ($M = 5.16, SD = 1.17$), which indicates that the

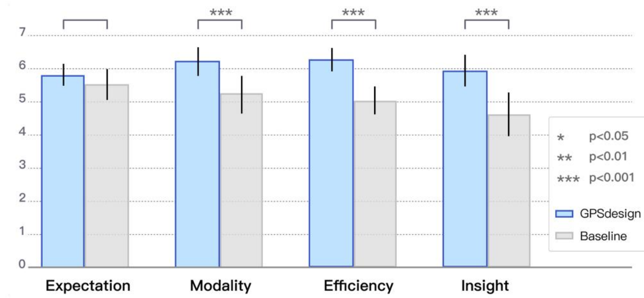

Figure 9. Comparison of self-customized questionnaire scores between GPSdesign and the baseline.

multi-modal information generated by GPSdesign better met participants' needs. For instance, image content could concretize the appearance of products (P2, P4, P8, P12), and the combination of images and text could promote understanding (P2, P16). Nevertheless, interviews reveal that the image content in both systems did not fully meet participants' expectations. Although the images generated by AI look very exquisite, they contain far less information than text (P1, P10, P12, P13, P14, P16). For example, P13 mentioned, *"I feel that text contains more information and provides more creativity. Some of the image generations are satisfactory to me, but they don't seem to bring as much creativity as text."* Some participants further pointed out that images cannot convey functional information and only provide appearance references (P1, P6, P15). Moreover, images are more prone to illusions than text. However, image illusions are not always detrimental; sometimes, incorrect information can actually inspire participants' creativity (P1, P11, P14), as P11 mentioned, *"I wasn't originally thinking of an indoor vertical farm scenario, but it sparked my imagination about indoor farming."*

For the efficiency dimension, the scores for GPSdesign ($M = 6.16$, $SD = 0.75$) are significantly higher than those for the baseline ($M = 4.94$, $SD = 0.85$). There are two main reasons for this outcome. On one hand, GPSdesign reduces the burden of writing prompts for participants (P2, P4, P7), and integrates AI content directly into the design network. For example, P2 mentioned, "*GPSdesign eliminates the step of writing prompts for me, where the previous node can serve as the basis for generating the next one, and it can maintain a high degree of relevance*," while using the baseline, participants (P1, P6) found it difficult to communicate with AI in unfamiliar domains. On the other hand, in GPSdesign, the AI responds quickly and concisely, serving as a point of reference and engaging in trial-and-error processes. This enables participants to focus solely on filtering feasible ideas, thereby augmenting their design efficiency (P1, P5, P6, P7, P9, P14). Conversely, in the control group, participants require additional time to formulate prompts, and the protracted AI-generated responses led to an increased cognitive load (P8, P14).

We also analyzed the insight dimension, and the results reveal that GPSdesign ($M = 5.81$, $SD = 0.93$) exhibited a statistically significant increase compared to the baseline ($M = 4.53$, $SD = 1.34$). Participants perceived that the co-evolutionary pattern in GPSdesign could stimulate profound thinking, leading to valuable insights (P1, P6, P7, P8). For the expectation dimension, the scores for GPSdesign ($M = 5.75$, $SD = 0.68$) do not significantly differ from those for the baseline ($M = 5.47$, $SD = 0.94$). This indicates that the information generated by AI in both systems generally aligned with participants' expectations.

7.2. RQ2: GPSdesign supports creativity effectively during conceptual design

To answer RQ2, we examined the CSI and the results of the design process coding. The content of the interviews served as supplementary explanations.

The statistical findings for the CSI are as follows: The paired sample t-test reveals that the overall score of GPSdesign ($M = 81.78, SD = 7.38$) was significantly higher than that of the baseline ($M = 66.28, SD = 8.89$), with a significance level of $p < 0.001$. This result indicates that, in the conceptual design process, GPSdesign provides greater creative support for designers. The distribution of scores across different dimensions for GPSdesign and baseline is shown in Figure 10. Since the data in each dimension satisfies a normal distribution, we also conducted a paired t-test for each dimension. The results indicate that, except for the immersion dimension, the average scores of GPSdesign are significantly higher than those of the baseline in all other dimensions.

Although in the immersion dimension, the average score of GPSdesign ($M = 6.47$) is higher than that of the baseline ($M = 6.06$), this difference is not significant. Some participants feel that the use of baseline reduces the sense of immersion (P1, P6, P13). This is attributed to the need for communication with GPT-4 during the design process, as P13 mentioned that *"I need to pause my thinking, and then figure out how to communicate with GPT-4, which feels like my train of thought is being interrupted."* Other participants believe that the immersive experience of using GPSdesign is inferior to that of baseline, as baseline tends to keep participants more focused (P3, P4, P5, P14). Specifically, P4 mentioned, *"In this system (baseline), everything needs to be thought out by oneself, whether one is thinking of a plan or asking AI, the collection of this information has to be done on one's own."*

The expressiveness dimension ($p < 0.001^{***}$) and results with effort dimension ($p = 0.002^{**}$) indicate that GPSdesign significantly enhances participants' efficiency in accomplishing design tasks compared to the baseline. The enjoyment dimension ($p < 0.001^{***}$) suggests that participants experienced greater creative satisfaction when using GPSdesign as opposed to the baseline. For the exploration dimension ($p < 0.001^{***}$), although the AI-generated information in both GPSdesign and baseline met participants' expectations, the diverse information in GPSdesign better supports participants' divergent thinking (P1, P5, P8, P12, P16). For example, P16 mentioned that *"because GPSdesign can diverge from one node to three nodes, this divergent pattern is slightly better than the one-to-one pattern of the baseline, as it reduces fixed thinking on a certain node."* Although not all three AI nodes provided by GPSdesign may be satisfactory each time, each node's content is as diverse as possible, which encourages participants to think from different perspectives, thereby continuously refining their ideas.

For the transformational actions that occur during the design process, we encode the nodes in the design space in the order they are added. The results of the design process coding are shown in Supplementary Appendix C.3. Overall, the number of effective nodes obtained by participants using GPSdesign was approximately twice that of using the baseline. In the early stages of design, participants add more problem nodes, while in the later stages of design, participants add more solution nodes. Participants typically spent more time alternating between problems and solutions when using GPSdesign compared to baseline. There are probably two reasons for the difference. One is that the co-evolutionary pattern in GPSdesign reminds participants to simultaneously advance the problem and solution. Another possible reason is that GPSdesign can help participants express the problems they are subconsciously thinking about, allowing the evolution of the problem to be structurally presented in the design space. For example, P9 mentioned, *"I actually kept my thoughts about the problem in my own mind and didn't record them, but in reality, this caused me to overlook many problems because I would forget the previous problems as I progressed. However, GPSdesign itself distinguishes between problem nodes and solution nodes, which reminded me to record them in a timely manner."*

Additionally, through the interviews, we learned that participants felt that using GPSdesign enabled them to generate ideas more quickly and fluidly while achieving higher-quality results. Firstly, the nodes in GPSdesign remind participants to advance both problems and solutions simultaneously, reducing the generation of ineffective questions and solutions. For example, P2 mentioned, "*GPSdesign guides me to consciously think about the problem and then the solution, or the solution and then look for the problem, like a ping-pong game with a back and forth,*" and this view is shared by P5, P9, P14. Additionally, the function of AI evaluation can further promote their reflection, thereby achieving higher quality design outcomes. Some participants find the potential match results useful, as P2 said, "*I see some dashed lines going from this node to that one, which might catch my attention, wondering if these two nodes are really related.*" Some participants believe that the matching results make the design ideas more visualizable, for example, P4 stated that "*It seems to allow me to see some issues I'm repeating or digging deeper into, which I hadn't realized I was exploring all along. Then, if I use AI evaluation earlier, it could remind me that I've been delving into this issue for a long time and should consider other issues.*" The vast majority of participants always assume that each generated node matches, but it is only after seeing the AI's evaluation results that they truly start to consider whether these nodes actually match or not.

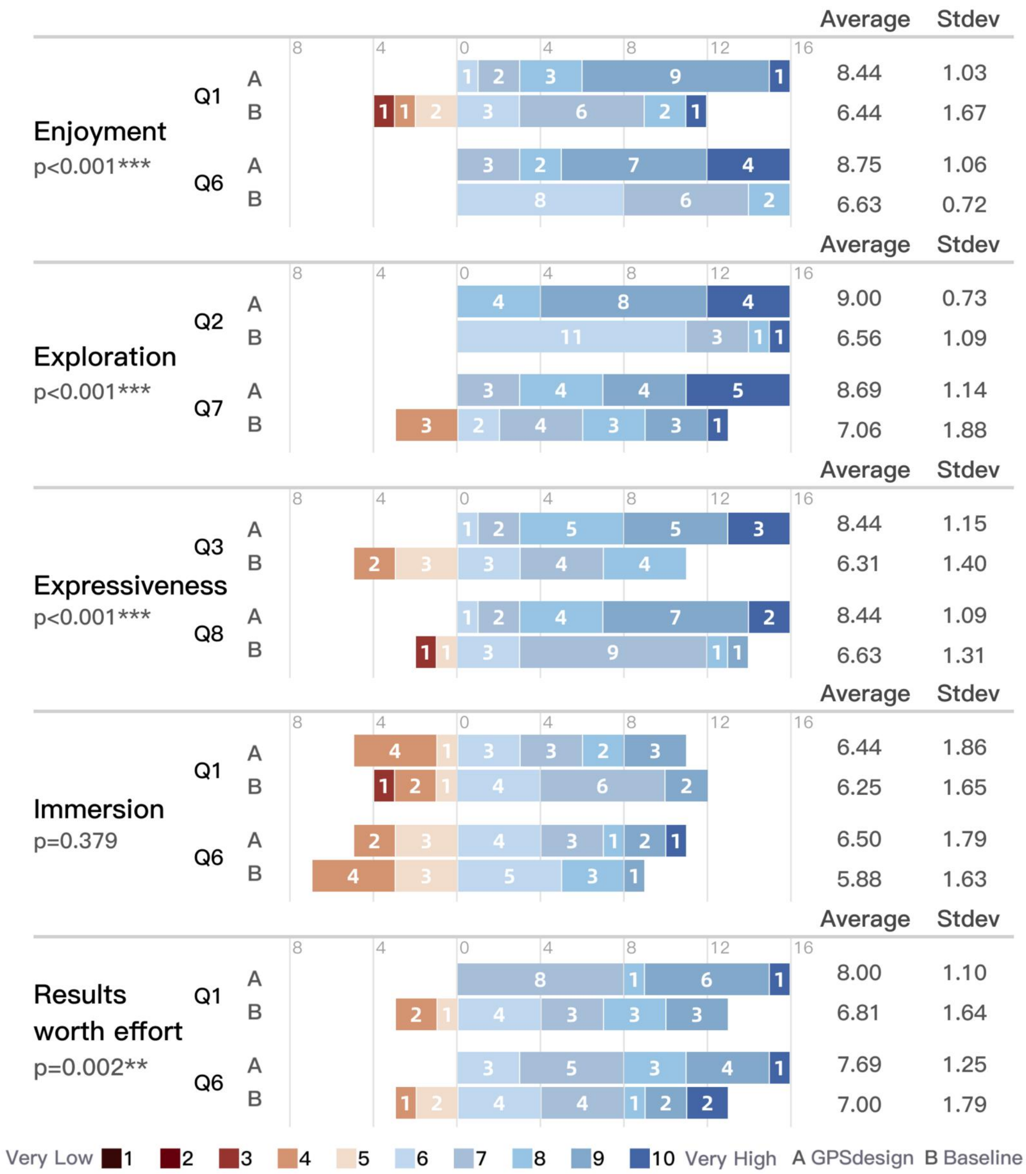

Figure 10. Comparison of CSI scale scores between GPSdesign and the baseline.

7.3. RQ3: GPSdesign is able to support different design thinking patterns

According to RQ1 and RQ2, GPSdesign made the user design process more efficient and effortless. In GPSdesign, the designer and AI produce content confined within the space of problem nodes and solution nodes. In the baseline, although the designer has more free control over the AI, the designer may lack balanced consideration of the problem space and the solution space. To further compare the impact of the co-evolutionary pattern on designers' creative thinking, we examined the coding results of the radar-chart-like visualization 6.3. From the coding results, we identified four dimensions of thinking patterns, namely the tendency towards divergence or convergence, the preference for alternation between problems and solutions, the preference for focusing on problems or solutions, and the sequence of AI involvement.

7.3.1. Divergence and convergence

As shown in Figure 11(a), P5 and P6 represent convergent design thinking and divergent design thinking, respectively. The total number of nodes produced by P5 and P6 is basically the same, but there is a significant difference in the distribution of the design space. On the one hand, networks with convergent design features are more focused on a particular direction in the circular distribution and have more iterative circular layers. This indicates that the participants tend to explore a key issue, and iteratively generate solutions or delve deeper into the problem starting from that issue. On the other hand, networks with divergent design features are more scattered and evenly distributed in the circular layout, with relatively fewer iterative circular layers. This indicates that the participants tend to comprehensively describe a problem and explore the optimal solutions for various aspects of that problem one by one.

7.3.2. Alternation between problems and solutions

As shown in Figure 11(b), P4 and P2 represent participants who are not accustomed to alternating between the problem and solution spaces, and participants who are accustomed to directly switching between the problem and solution spaces, respectively. Participants who do not frequently switch between the problem and solution spaces tend to focus solely on either the problem space or the solution space until they determine that there is nothing more to add to the current space before considering the other side. With this mindset, participants may easily get stuck on a particular aspect. One coping strategy is to use the evaluation function in a timely manner. As P4 mentioned, *"The results of the evaluation stage allowed me to see some issues that I was repeating or digging deeper into, which I was not aware of being stuck on before. For example, here I was constantly involved in the design of the robotic arm base. If I had used the evaluation function earlier, it might have reminded me that I had been delving into this issue for a long time and should consider other issues."* Participants who are adept at frequently switching between the problem space and the solution space tend to be skilled at discovering new problems from solutions. This way of thinking may be conducive to promoting the feasibility of the final solution. For example, P7 mentioned, *"Through this back-and-forth process from S to P and P to S, every solution I came up with not only stayed at the conceptual level but also included how to execute and implement the concept, eventually leading to a more detailed and feasible solution."* However, frequently switching between the problem and solution spaces may cause participants' thinking to become disorganized. For instance, P13 mentioned, *"Although I came up with new ideas or problems, this might have interrupted my original train of thought, and I would forget what the main focus was."*

7.3.3. The focus on problems or solutions

As shown in Figure 11(c), P8 and P9 represent participants who focus on exploring the problem space and the solution space, respectively. The most obvious difference between them lies in the exploration of the initial problem space and the overall number of problem nodes and solution nodes. The problem-oriented design thinking makes participants more cautious in defining the initial problem, while the solution-oriented design thinking focuses more on exploring the diversity and feasibility of solutions. However, the designer's personal habits are not the only reason that leads to the imbalance between the problem and solution spaces. Whether AI can effectively provide information will also affect the designer's efficiency in exploring the current space. This point has already been discussed in section 7.1. P8 had a more open and inclusive attitude towards using AI for problem and requirement exploration, and he said *"At the beginning, I wasn't very familiar with the agricultural field and robotic arms, so I would choose to let AI first generate some potential problems it thinks of, and then add some problems that I think are meaningful. Without AI, I might need to do a lot of research."* P9 believes that AI is more adept at proposing specific functions and solutions; therefore, he had a broader exploration of the solution space.

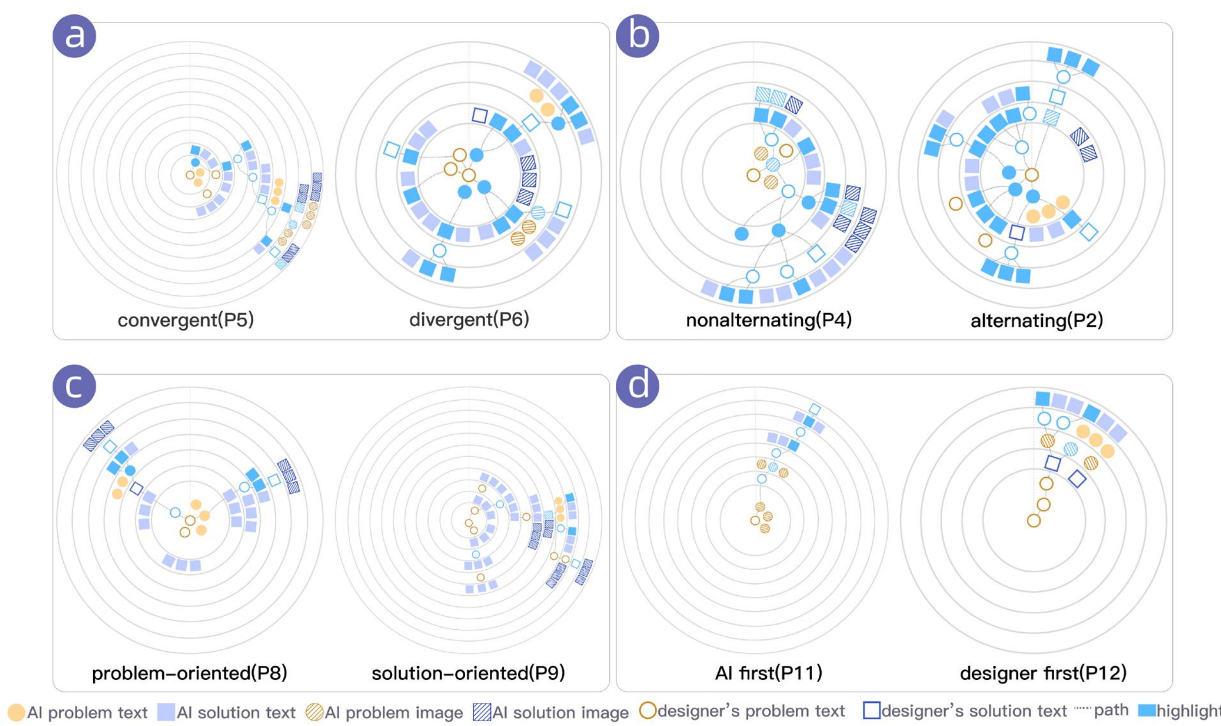

Figure 11. Dimensions of designers' thinking patterns during using GPSdesign. (a) Divergence and convergence; (b) Alternation between problems and solutions; (c) focus on problems or solutions; (d) sequence of AI involvement.

7.3.4. The sequence of AI involvement

As shown in Figure 11(d), P11 represents participants who are accustomed to having AI generate first, and then think for themselves; P12 represents participants who are accustomed to thinking for themselves first, and then using AI to generate. For the former, AI acts as a suggester, while the designer is responsible for reviewing and filtering. For example, P11 mentioned, *"I need AI to stimulate some of my inspirations or help me diverge into different scenarios, and then I consider and converge on this content."* However, this may cause the designer to overly rely on AI and lack their own thinking (Ma et al., 2024). Consistent with previous research, P11 suggests that AI impeded his ability to think independently. As for the latter, the designer proposes the main ideas, and AI supplements the content to make the design more comprehensive and complete. For instance, P12 stated that *"AI can help me fill in the missing cases to make it complete, and then I consider whether to adopt this content."*

8. Discussion

Our research integrates GenAI capabilities with the problem-solution co-evolution model. In the user study, we examined how the GPSdesign system supports designers throughout the design process and its influence on designers' thinking patterns. The results demonstrate that GenAI embedded within the co-evolution model significantly alters the way designers develop their ideas. Furthermore, the proposed and developed GPSdesign system provides a practical example of integrating design research theory with GenAI, demonstrating the feasibility and application value of design research theory in human–GenAI collaborative design.

8.1. Enhancing design space expansion using GPSdesign

Designers using the GPSdesign system experience enhanced support for creative expression, efficiency, and quality during the design process compared to a baseline system. This improvement may come from the "shared design context" between GenAI and the designers in the GPSdesign system. The "shared design context" refers to that GPSdesign integrates information directly into the design space, preserving the causal relationship of design nodes and allowing both designers and AI to reference and build upon any previous node in the process. This approach aligns with recent developments in AI-assisted design tools, such as CausalMapper (Huang et al., 2023) and AI-Assisted Causal Pathway Diagram (Zhong et al., 2024), which similarly minimize the time spent creating visualizations while fostering GenAI to understand causal relationship. In contrast, while the baseline system gives designers more control over GenAI, it introduces challenges in efficiently integrating AI-generated information into their designs and constructing valuable references.

Furthermore, the evaluation function in the GPSdesign system also promotes designers' reflection on their design space. It helps them identify connections between design problems and solutions, thereby producing higher-quality design outcomes. Some designers have indicated that the evaluation results have made them aware of certain issues they have been exploring or have helped them trace their thought processes, which is beneficial for improving design quality.

8.2. Balancing problem and solution spaces through GPSdesign

Through the encoding analysis of the design space network, we found that different designers exhibit distinct stylistic characteristics when using the GPSdesign system. For example, some designers prefer to frequently switch between problems and solutions; others solely focus on exploring one side. These differences primarily stem from the designers' cognitive styles and design experiences.

However, the analysis revealed that designers utilizing the GPSdesign system demonstrate a notably increased frequency of transitioning between problems and solutions throughout the design process compared to those using the baseline system. This phenomenon may be attributed to the co-evolution model employed in GPSdesign, which prompts designers to concurrently advance their exploration of both problems and solutions rather than concentrating solely on one aspect. Additionally, the GPSdesign system helps designers to better express and organize the design problems in their subconscious, thereby facilitating a clearer evolution of the problem space within the design space.

8.3. Limitations and future work

Although the results of the user study indicate the benefits of the GPSdesign, there are still some limitations that need further improvement.

Firstly, this study only focuses on the multi-modal information provided to designers but does not delve into whether GenAI can support automated co-evolutionary logical reasoning. Although GenAI's reasoning ability is insufficient to handle the product design process (Alam et al., 2024), research has shown that prompting technologies such as chain of thought (CoT) can improve logical reasoning ability (Feng et al., 2024). Future research could further explore how GenAI supports reasoning in co-evolutionary design.

Secondly, there are doubts about whether GenAI possesses sufficient expertise. Although most models are trained on extensive data [including a comprehensive knowledge base in the fields of design and manufacturing (Makatura et al., 2023)], it is unknown whether they have expert-level knowledge of product and engineering design. For example, GenAI suggested P15 to "use carbon fiber composites to build drones," but P15 was confused about whether the suggestion was feasible and whether there was a better alternative. This requires an excessive amount of effort for certification. Besides, the presence of hallucinations further reduces the designer's confidence in the generative information, especially when the design is very complex (Leiser et al., 2023; Perković et al., 2024). Future research can further explore how to enhance GenAI's ability to understand, interpret, and apply professional design knowledge.

Thirdly, regarding the construction of a structured design space, some participants mentioned that GPSdesign only supports the addition of nodes, not the merging and convergence of nodes. Future studies could explore the principles of node operations in the co-evolutionary space of problems and solutions to support more flexible construction of structured design spaces. Finally, based on the different styles of using the co-evolution model found in RQ3, future research could also explore how to provide more personalized design support based on different designers' cognitive styles.

There are also some methodological limitations in the formative research experiment and the user research experiment adopted in this study. In our study, the participants were mainly seniors majoring in industrial design and mechanical design. Although we believe that they have at least three years of practical training experience in design, they still lack practical experience in real environments and a deep understanding of the conceptual design process of products. Since one of the motivations for this study was to improve designers' creativity without disrupting the workflow, relying on student experiments alone may not adequately explore this issue. Future research needs to call on more professional designers to participate to obtain a deeper and more realistic insight. In addition, we only recruited 16 participants for the user study, which may not be sufficient to fully explore and classify designers' thinking patterns, so further experiments need to increase the number of participants.

9. Conclusion

This article proposes a generative problem-solution co-evolution method and developed a system GPSdesign, accordingly. The proposed method facilitates the conceptual design process by organizing diverse design content, utilizing GenAI to expand problem and solution spaces, and assisting designers in identifying optimal solutions within intricate design spaces. Experimental results show that the SUS scores and CSI scores of the GPSdesign system are significantly higher than those of the baseline. Within the co-evolutionary pattern, designers generate more balanced problem and solution spaces, leading to enhanced performance in human–AI collaborative design. The study also identifies variations in designers' thinking patterns when employing the co-evolution model, prompting further exploration into promoting shared design contexts between GenAI and designers to enhance the co-evolutionary methodology of human–GenAI collaboration.

Notes

1. https://openai.com/index/gpt-4
2. https://www.anthropic.com/news/claude-3-5-sonnet
3. https://stability.ai/stable-image
4. https://openai.com/index/dall-e-3
5. https://chatgpt.com
6. https://www.midjourney.com
7. https://stablediffusionweb.com
8. https://flask.palletsprojects.com
9. https://react.dev
10. https://www.mongodb.com
11. https://openai.com/index/gpt-4
12. https://openai.com/index/dall-e-3

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Key R & D Program of China (2022YFB3303301).

ORCID

Pei Chen http://orcid.org/0000-0003-0962-6459
Lingyun Sun http://orcid.org/0000-0002-5561-0493

References

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., Bello, I., … Szym. (2023). GPT-4 technical report. *arXiv Preprint*, arXiv:2303.08774.

Alam, M. F., Lentsch, A., Yu, N., Barmack, S., Kim, S., Acemoglu, D., Hart, J., Johnson, S., & Ahmed, F. (2024). From automation to augmentation: Redefining engineering design and manufacturing in the age of NextGen-AI. *An MIT Exploration of Generative AI*. https://doi.org/10.21428/e4baedd9.e39b392d

Benami, O., & Jin, Y. (2002). Creative stimulation in conceptual design. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (Vol. 3624, pp. 251–263). https://doi.org/10.1115/DETC2002/DTM-34023

Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability Evaluation in Industry* (pp. 189–194). Taylor & Francis.

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). *Generative AI at work*. Technical report. National Bureau of Economic Research.

Carroll, E. A., Latulipe, C., Fung, R., & Terry, M. (2009). Creativity factor evaluation: Towards a standardized survey metric for creativity support. In *Proceedings of the Seventh ACM Conference on Creativity and Cognition* (pp. 127–136).

Cash, P., Gonçalves, M., & Dorst, K. (2023). Method in their madness: Explaining how designers think and act through the cognitive co-evolution model. *Design Studies*, *88*, 101219. https://doi.org/10.1016/j.destud.2023.101219

Casteleiro-Pitrez, J. (2024). Generative artificial intelligence image tools among future designers: A usability, user experience, and emotional analysis. *Digital*, *4*(2), 316–332. https://doi.org/10.3390/digital4020016

Chen, C., Nguyen, C., Groueix, T., Kim, V. G., & Weibel, N. (2024). Memovis: A GenAI-powered tool for creating companion reference images for 3d design feedback. *arXiv Preprint*, arXiv:2409.06082.

Cherry, E., & Latulipe, C. (2014). Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer–Human Interaction*, *21*(4), 1–25. https://doi.org/10.1145/2617588

Chung, J. J. Y., & Adar, E. (2023). Promptpaint: Steering text-to-image generation through paint medium-like interactions. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (pp. 1–17). https://doi.org/10.1145/3586183.3606777

Crilly, N. (2021a). The evolution of "co-evolution" (part I): Problem solving, problem finding, and their interaction in design and other creative practices. *She Ji: The Journal of Design, Economics, and Innovation*, *7*(3), 309–332. https://doi.org/10.1016/j.sheji.2021.07.003

Crilly, N. (2021b). The evolution of "co-evolution" (part II): The biological analogy, different kinds of co-evolution, and proposals for conceptual expansion. *She Ji: The Journal of Design, Economics, and Innovation*, *7*(3), 333–355. https://doi.org/10.1016/j.sheji.2021.07.004

Crilly, N., & Firth, R. M. (2019). Creativity and fixation in the real world: Three case studies of invention, design and innovation. *Design Studies*, *64*, 169–212. https://doi.org/10.1016/j.destud.2019.07.003

Daalhuizen, J., Badke-Schaub, P., & Batill, S. (2009). Dealing with uncertainty in design practice: Issues for designer-centered methodology. In *17th International Conference on Engineering Design* (pp. 147–158). Design Society.

Demirel, H. O., Goldstein, M. H., Li, X., & Sha, Z. (2024). Human-centered generative design framework: An early design framework to support concept creation and evaluation. *International Journal of Human–Computer Interaction*, *40*(4), 933–944. https://doi.org/10.1080/10447318.2023.2171489

Dorst, K. (2004). On the problem of design problems-problem solving and design expertise. *Journal of Design Research*, *4*(2), 0. https://doi.org/10.1504/JDR.2004.009841

Dorst, K. (2019). Co-evolution and emergence in design. *Design Studies*, *65*, 60–77. https://doi.org/10.1016/j.destud.2019.10.005

Dorst, K., & Cross, N. (2001). Creativity in the design process: Co-evolution of problem–solution. *Design Studies*, *22*(5), 425–437. https://doi.org/10.1016/S0142-694X(01)00009-6

Dow, S. P., Glassco, A., Kass, J., Schwarz, M., Schwartz, D. L., & Klemmer, S. R. (2010). Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. *ACM Transactions on Computer–Human Interaction*, *17*(4), 1–24. https://doi.org/10.1145/1879831.1879836

Farrell, R., & Hooker, C. (2013). Design, science and wicked problems. *Design Studies*, *34*(6), 681–705. https://doi.org/10.1016/j.destud.2013.05.001

Feng, G., Zhang, B., Gu, Y., Ye, H., He, D., & Wang, L. (2024). Towards revealing the mystery behind chain of thought: A theoretical perspective. In *Proceedings of the 37th International Conference on Neural Information Processing Systems* (p. 42). Curran Associates Inc.

Fiorineschi, L., Rotini, F., & Rissone, P. (2016). A new conceptual design approach for overcoming the flaws of functional decomposition and morphology. *Journal of Engineering Design*, *27*(7), 438–468. https://doi.org/10.1080/09544828.2016.1160275

Garvey, B., & Childs, P. (2015). Design as an unstructured problem: New methods to help reduce uncertainty—A practitioner perspective. In *Impact of design research on industrial practice: Tools, technology, and training* (pp. 333–352). Springer.

Gero, J. S., Kannengiesser, U., & Crilly, N. (2022). Abstracting and formalising the design co-evolution model. *Design Science*, *8*, e14. https://doi.org/10.1017/dsj.2022.10

Guo, J., Yin, Y., Sun, L., & Chen, L. (2024). Empirical study of problem-solution co-evolution in Human–GAI collaborative conceptual design. DRS2024, Boston, MA, USA. https://doi.org/10.21606/drs.2024.983

Haase, J., Djurica, D., & Mendling, J. (2023). The art of inspiring creativity: Exploring the unique impact of AI-generated images. In *AMCIS 2023 Proceedings, Number 10 in Special Interest Group on Artificial Intelligence and Autonomous Applications*.

Haynes, P., & Yang, S. (2023). Supersystem digital twin-driven framework for new product conceptual design. *Advanced Engineering Informatics*, *58*, 102149. https://doi.org/10.1016/j.aei.2023.102149

He, Q., Zheng, W., Bao, H., Chen, R., & Tong, X. (2023). Exploring designers' perceptions and practices of collaborating with generative AI as a co-creative agent in a multi-stakeholder design process: Take the domain of avatar design as an example. In *Proceedings of the Eleventh International Symposium of Chinese CHI* (pp. 596–613). https://doi.org/10.1145/3629606.3629675

Hong, M. K., Hakimi, S., Chen, Y.-Y., Toyoda, H., Wu, C., & Klenk, M. (2023). Generative AI for product design: Getting the right design and the design right. *arXiv Preprint*, arXiv:2306.01217.

Huang, Z., Quan, K., Chan, J., & MacNeil, S. (2023). Causalmapper: Challenging designers to think in systems with causal maps and large language model. In *Proceedings of the 15th Conference on Creativity and Cognition* (pp. 325–329). https://doi.org/10.1145/3591196.3596818

Hui, Q., Li, Y., Tao, Y., & Liu, H. (2020). Triple-helix structured model based on problem-knowledge-solution co-evolution for innovative product design process. *Chinese Journal of Mechanical Engineering*, *33*(1), 94. https://doi.org/10.1186/s10033-020-00519-2

Hwang, D., & Park, W. (2018). Design heuristics set for x: A design aid for assistive product concept generation. *Design Studies*, *58*, 89–126. https://doi.org/10.1016/j.destud.2018.04.003

Ikoma, D., Aoki, E., Taniguchi, T., Suzuki, S., & Ohkuma, T. (2024). Automatic design summary generation with generative AI. Companion Proceedings of the ACM on Web Conference 2024 (pp. 1313–1317). https://doi.org/10.1145/3589335.3651901

Jiang, J. (2024). When generative artificial intelligence meets multimodal composition: Rethinking the composition process through an AI-assisted design project. *Computers and Composition*, *74*, 102883. https://doi.org/10.1016/j.compcom.2024.102883

Kim, E., & Kim, K. (2015). Cognitive styles in design problem solving: Insights from network-based cognitive maps. *Design Studies*, *40*, 1–38. https://doi.org/10.1016/j.destud.2015.05.002

Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open*, *6*, 100174. https://doi.org/10.1016/j.caeo.2024.100174

Lee, M. (2023). A mathematical investigation of hallucination and creativity in GPT models. *Mathematics*, *11*(10), 2320. https://doi.org/10.3390/math11102320

Leiser, F., Eckhardt, S., Knaeble, M., Maedche, A., Schwabe, G., & Sunyaev, A. (2023). From ChatGPT to FactGPT: A participatory design study to mitigate the effects of large language model hallucinations on users. In *Proceedings of Mensch Und Computer 2023* (pp. 81–90). https://doi.org/10.1145/3603555.3603565

Li, W., Li, Y., Wang, J., & Liu, X. (2010). The process model to aid innovation of products conceptual design. *Expert Systems with Applications*, *37*(5), 3574–3587. https://doi.org/10.1016/j.eswa.2009.10.034

Liu, F., Lv, J., Cui, S., Luan, Z., Wu, K., & Zhou, T. (2024). "Smart error"! Exploring imperfect AI to support creative ideation. *Proceedings of the ACM on Human–Computer Interaction*, *8*(CSCW1), 1–28. https://doi.org/10.1145/3637398

Liu, H., Zhang, X., Zhou, J., Shou, Y., Yin, Y., & Chai, C. (2024). Cognitive styles and design performances in conceptual design collaboration with GenAI. *International Journal of Technology and Design Education*. https://doi.org/10.1007/s10798-024-09937-y

Ma, K., Moore, G., Shyam, V., Villarrubia, J., Goucher-Lambert, K., & Reynolds Brubaker, E. (2024). Human–AI collaboration among engineering and design professionals: Three strategies of generative AI use. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (Vol. 88407, p. V006T06A025). American Society of Mechanical Engineers. https://doi.org/10.1115/DETC2024-143560

Maher, M. L. (1990). Process models for design synthesis. *AI Magazine*, *11*(4), 49–49. https://doi.org/10.1609/aimag.v11i4.856

Maher, M. L., & Wu, P. X. (1999). Reconsidering fitness and convergence in co-evolutionary design. In *Australian Joint Conference on Artificial Intelligence* (pp. 488–489). Citeseer.

Maher, M. L., Poon, J., & Boulanger, S. (1996). Formalising design exploration as co-evolution: A combined gene approach. In *Advances in Formal Design Methods for CAD: Proceedings of the IFIP WG5. 2 Workshop on Formal Design Methods for Computer-Aided Design, June 1995* (pp. 3–30). Springer.

Maher, M., & Tang, H.-H. (2003). Co-evolution as a computational and cognitive model of design. *Research in Engineering Design*, *14*(1), 47–64. https://doi.org/10.1007/s00163-002-0016-y

Makatura, L., Foshey, M., Wang, B., Hähnlein, F., Ma, P., Deng, B., Tjandrasuwita, M., Spielberg, A., Owens, C. E., Chen, P. Y., Zhao, A., Zhu, A., Norton, W. J., Gu, E., Jacob, J., Li, Y., Schulz, A., & Matusik, W. (2023). How can large language models help humans in design and manufacturing? *arXiv Preprint*, arXiv:2307.14377.

Martinec, T., Škec, S., Perišić, M. M., & Štorga, M. (2020). Revisiting problem-solution co-evolution in the context of team conceptual design activity. *Applied Sciences*, *10*(18), 6303. https://doi.org/10.3390/app10186303

Mbalaka, B. (2023). Epistemically violent biases in artificial intelligence design: The case of DALLE-E 2 and starry AI. *Digital Transformation and Society*, *2*(4), 376–402. https://doi.org/10.1108/DTS-01-2023-0003

Morris, M. R., Cai, C. J., Holbrook, J., Kulkarni, C., & Terry, M. (2023). The design space of generative models kalving creativity. *arXiv Preprint*, arXiv:2304.10547.

Morse, E., Dantan, J.-Y., Anwer, N., Söderberg, R., Moroni, G., Qureshi, A., Jiang, X., & Mathieu, L. (2018). Tolerancing: Managing uncertainty from conceptual design to final product. *CIRP Annals*, *67*(2), 695–717. https://doi.org/10.1016/j.cirp.2018.05.009

Perković, G., Drobnjak, A., & Botički, I. (2024). Hallucinations in LLMS: Understanding and addressing challenges. 2024 47th MIPRO ICT and Electronics Convention (MIPRO) (pp. 2084–2088). IEEE. https://doi.org/10.1109/MIPRO60963.2024.10569238

Rasal, S., & Hauer, E. (2024). Navigating complexity: Orchestrated problem solving with multi-agent LLMS. *arXiv Preprint*, arXiv:2402.16713.

Saadi, J. I., & Yang, M. C. (2023). Generative design: Reframing the role of the designer in early-stage design process. *Journal of Mechanical Design*, *145*(4), 041411. https://doi.org/10.1115/1.4056799

Sarkar, P., & Chakrabarti, A. (2011). Assessing design creativity. *Design Studies*, *32*(4), 348–383. https://doi.org/10.1016/j.destud.2011.01.002

Schön, D. A. (1992). *The reflective practitioner: How professionals think in action* (1st ed.). Routledge.

Shen, Y., Shen, Y., Cheng, J., Jiang, C., Fan, M., & Wang, Z. (2024). Neural canvas: Supporting scenic design prototyping by integrating 3d sketching and generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1–18). https://doi.org/10.1145/3613904.3642096

Simkute, A., Tankelevitch, L., Kewenig, V., Scott, A. E., Sellen, A., & Rintel, S. (2024). Ironies of generative AI: Understanding and mitigating productivity loss in human–AI interactions. *International Journal of Human–Computer Interaction*, 1–22. https://doi.org/10.1080/10447318.2024.2405782

Smits, A., & van Turnhout, K. (2023). A critical curation of solution repertoire by first time design students. Proceedings of the International Conference on Engineering and Product Design Education E&PDE, Utrecht, Netherlands.

Smulders, F., Reyman, I., & Dorst, K. (2009). Modelling co-evolution in design practice. In *Proceedings of the 17th International Conference on Engineering Design, ICED* (pp. 2–335–2–346). The Design Society.

Sun, Y., Jang, E., Ma, F., & Wang, T. (2024). Generative AI in the wild: Prospects, challenges, and strategies. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1–16). https://doi.org/10.1145/3613904.3642160

Tankelevitch, L., Kewenig, V., Simkute, A., Scott, A. E., Sarkar, A., Sellen, A., & Rintel, S. (2023). The metacognitive demands and opportunities of generative AI. *arXiv Preprint*, arXiv:2312.10893.

Tembhekar, P., Devan, M., & Jeyaraman, J. (2023). Role of GenAI in automated code generation within DevOps practices: Explore how generative AI. *Journal of Knowledge Learning and Science Technology*, *2*(2), 500–512. https://doi.org/10.60087/jklst.vol2.n2.p512

Wadinambiarachchi, S., Kelly, R. M., Pareek, S., Zhou, Q., & Velloso, E. (2024). The effects of generative AI on design fixation and divergent thinking. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1–18). https://doi.org/10.1145/3613904.3642919

Wang, D., & Han, J. (2023). Exploring the impact of generative stimuli on the creativity of designers in combinational design. In *Proceedings of the Design Society* (Vol. 3, pp. 1805–1814).

Wang, S.-Y., Su, W.-C., Chen, S., Tsai, C.-Y., Misztal, M., Cheng, K. M., Lin, A., Chen, Y., & Chen, M. Y. (2024). Roomdreaming: Generative-AI approach to facilitating iterative, preliminary interior design exploration. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1–20). https://doi.org/10.1145/3613904.3642901

Wang, Y., Damen, N. B., Gale, T., Seo, V., & Shayani, H. (2024). Inspired by AI? a novel generative AI system to assist conceptual automotive design. *arXiv Preprint*, arXiv:2407.11991. https://doi.org/10.1115/DETC2024-142124

Wiltschnig, S., Christensen, B. T., & Ball, L. J. (2013). Collaborative problem–solution co-evolution in creative design. *Design Studies*, *34*(5), 515–542. https://doi.org/10.1016/j.destud.2013.01.002

Winske, D., & Omidi, N. (1991). Hybrid codes: Methods and applications. Presented at the 4th International School for Space Simulation (pp. 1–5).

Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining whether, why, and how human–AI interaction is uniquely difficult to design. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (pp. 1–13). https://doi.org/10.1145/3313831.3376301

Yao, J., Chen, P., Li, Z., Cai, Y., Wu, Y., You, W., & Sun, L. (2024). StepIdeator: Utilizing mixed representations to support step-by-step design with generative AI. *Journal of Mechanical Design*, 1–78. https://doi.org/10.1115/1.4067426

Yu, K., Xiao, Y., Li, M., Yu, S., Yang, Y., Guo, X., Zhang, W., & Yuan, X. (2024). Design for AI-integrated design team collaboration: A strategy and exploration using node flow in establishing a reusable representation of knowledge in the collaborative process. DRS2024, Boston, MA, USA. https://doi.org/10.21606/drs.2024.985

Yu, R., Gu, N., Ostwald, M., & Gero, J. S. (2015). Empirical support for problem–solution coevolution in a parametric design environment. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, *29*(1), 33–44. https://doi.org/10.1017/S0890060414000316

Yüksel, N., Börklü, H. R., Sezer, H. K., & Canyurt, O. E. (2023). Review of artificial intelligence applications in engineering design perspective. *Engineering Applications of Artificial Intelligence*, *118*, 105697. https://doi.org/10.1016/j.engappai.2022.105697

Zhang, H., Chen, P., Xie, X., Lin, C., Liu, L., Li, Z., You, W., & Sun, L. (2024). Protodreamer: A mixed-prototype tool combining physical model and generative AI to support conceptual design. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (pp. 1–18). https://doi.org/10.1145/3654777.3676399

Zhang, K., Cai, S., Yang, W., Wu, W., & Shen, H. (2024). Exploring optimal combinations: The impact of sequential multimodal inspirational stimuli in design concepts on creativity. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (pp. 2788–2801). https://doi.org/10.1145/3643834.3661501

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., … Wen, J.-R. (2023). A survey of large language models. *arXiv Preprint*, arXiv:2303.18223.

Zhong, R., Shin, D., Meza, R., Klasnja, P., Colusso, L., & Hsieh, G. (2024). AI-assisted causal pathway diagram for human-centered design. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1–19). https://doi.org/10.1145/3613904.3642179

Zhou, J., Li, R., Tang, J., Tang, T., Li, H., Cui, W., & Wu, Y. (2024). Understanding nonlinear collaboration between human and AI agents: A co-design framework for creative design. *arXiv Preprint*, arXiv:2401.07312.

About the authors

Pei Chen is a postdoctoral research fellow at the College of Computer Science and Technology, Zhejiang University. Her main

research interests are human–AI collaboration and intelligent design methods.

Yexinrui Wu is a master candidate at the College of Computer Science and Technology, Zhejiang University. Her research focuses on intelligent human–AI interaction systems, exploring innovative applications of artificial intelligence in design.

Zhoushu Li is a master graduate from the College of Computer Science and Technology, Zhejiang University. Her main research interests are human–computer interactions during design, human-centered design methods.

Hongbo Zhang is a PhD candidate at the College of Computer Science and Technology, Zhejiang University. His doctoral research focuses on human–AI interactions and exploring innovative methods and tools for AI-supported conceptual design.

Mingxu Zhou is a master candidate at the College of Computer Science and Technology, Zhejiang University. His research focuses on multimodal interaction design for generative models.

Jiayi Yao is a master candidate at the School of Software Technology, Zhejiang University. Her research focuses on developing innovative methods and tools for conceptual design supported by large language models.

Weitao You is a research assistant at the College of Computer Science and Technology, Zhejiang University. His research focuses on computational aesthetics, design intelligence, and human–computer interaction.

Lingyun Sun is a professor at the College of Computer Science and Technology, Zhejiang University. His research interests include design intelligence, innovation and design, as well as information and interaction design.