# CoExploreDS: Framing and Advancing Collaborative Design Space Exploration Between Human and AI

Pei Chen
College of Computer Science and Technology
Zhejiang University
Hangzhou, China
chenpei@zju.edu.cn

Jiayi Yao
School of Software Technology
Zhejiang University
Hangzhou, China
yao_jiayi@zju.edu.cn

Zhuoyi Cheng
School of Software Technology
Zhejiang University
Hangzhou, China
zy_cheng@zju.edu.cn

Yichen Cai
College of Computer Science and Technology
Zhejiang University
Hangzhou, China
yichencai@zju.edu.cn

Jiayang Li
College of Computer Science and Technology
Zhejiang University
Hangzhou, China
jiayangli@zju.edu.cn

Weitao You
College of Computer Science and Technology
Zhejiang University
Hangzhou, China
weitao_you@zju.edu.cn

Lingyun Sun*
College of Computer Science and Technology
Zhejiang University
Hangzhou, China
sunly@zju.edu.cn

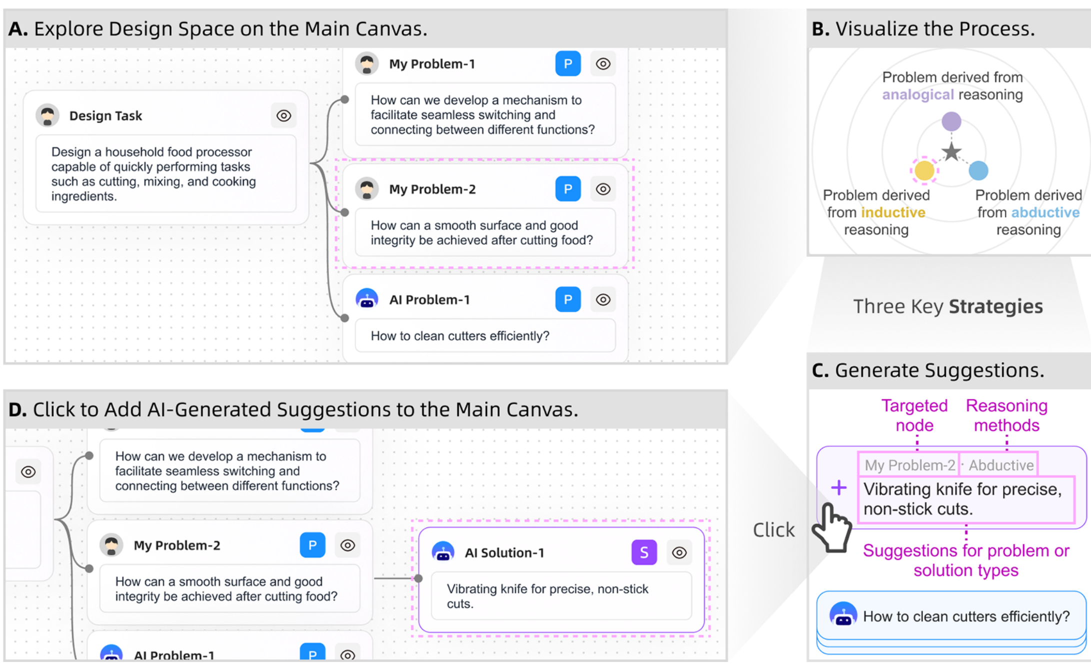


**Figure 1: A case illustration of designer-AI collaborative design space exploration in CoExploreDS. In this case, the designer began by (A) outlining the design task on the main canvas, focusing on the development of a household food processor. The designer focused on the problem, "*How can a smooth surface and good integrity be achieved after cutting food?*" (B) Using the problem-solution co-evolution model and design reasoning methods, CoExploreDS visualizes the design process, analyzes the current design process, and (C) generates multiple possible suggestions during the conceptualization process. Subsequently, the designer will be inspired by the suggestion "*Vibrating knife for precise, non-stick cuts*", and then (D) clicks to add this node to the main canvas, thereby exploring the design space together with AI.**

*Corresponding Author


*CHI '25, Yokohama, Japan*

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713869

## Abstract

In product design, effective design space exploration (DSE) is crucial for generating high-quality design ideas, requiring designers to possess broad knowledge and balance various constraints. As large-scale models thrive, AI has become an indispensable design collaborator by providing cross-domain knowledge and assistance with complex reasoning. To facilitate collaborative DSE between designers and AI, we frame and advance the design process through the problem-solution co-evolution model and design reasoning methods. A formative study was conducted to identify key strategies for the implementation. Then we developed CoExploreDS, a system that formalizes problems and solutions emerging in the human-AI collaborative design space into nodes. Using four reasoning methods, this system dynamically generates suggestions based on the ongoing design process. User studies confirmed that CoExploreDS significantly improves design quality and the human-AI collaboration experience.

## CCS Concepts

• **Human-centered computing → Interactive systems and tools**.

## Keywords

Human-AI collaboration, Design space exploration, Generative AI

**ACM Reference Format:**

Pei Chen, Jiayi Yao, Zhuoyi Cheng, Yichen Cai, Jiayang Li, Weitao You, and Lingyun Sun. 2025. CoExploreDS: Framing and Advancing Collaborative Design Space Exploration Between Human and AI. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3706598.3713869

## 1 Introduction

In the product design process, the quality of ideas generated in the initial phase is essential as it significantly impacts the development costs and the level of innovation in the final product [55, 66]. In this phase, designers strive to generate high-quality solutions by exploring a product's "design space" [74]. This process involves refining the initial design task into detailed requirements and exploring all possible and alternative solutions within practical constraints [66]. However, design space exploration (DSE) is demanding for designers because of its information-intensive nature [8, 32]. Designers must master a broad spectrum of knowledge and develop potential solutions comprehensively, thereby expanding the boundaries of the design space [4, 80]. Additionally, they navigate trade-offs and constraints to reason across various problems and solutions, transforming ill-defined problems into feasible solutions [56].

Previous theories on the early stages of design have provided valuable references and guidelines for understanding and managing the process of DSE. To frame information within design spaces, the problem-solution co-evolution model has been widely adopted and continuously developed [15, 23]. This model divides the design space into two parts: a "problem space" and a "solution space". As designers explore, design problems and solutions "co-evolve" over time until a good "matching" between problem and solution is achieved. Not only are solutions assessed within the context of the specific problem, but requirements can also be further adjusted based on new solutions [73].

To advance DSE, understanding the driving force underlying the evolution of problems and solutions under specific constraints is crucial. Designers may employ various methods, including deductive [14], inductive [22], abductive [20], and analogical reasoning [7, 13]. For example, assuming that *a portable device made of light composite materials will likely be popular if surveys support preferences for such materials* demonstrates inductive reasoning. Meanwhile, inferring that *a smartwatch's short battery life results from its always-on display because it consumes more power* exemplifies abductive reasoning. With these methods, designers can apply positive constraints or backward inferences [28, 47], thus addressing diverse considerations in complex design practice.

However, applying design theories in work predominantly relies on designers' years of training and intuition [31, 65]. These practices call for creative tools that support rapid ideation across various knowledge domains and structure the thinking process in alignment with design theories. As an alternative, large language models (LLMs) are increasingly employed in human-AI collaborative design to supplement knowledge [35] and assist reasoning [10] for their rich knowledge and strong in-context comprehension abilities. For example, some collaborative systems such as GenQuery [63] and DiscipLink [79] enable designers to utilize similarity-based search or automatic query expansion, thus aggregating diverse results from interdisciplinary sources into the design space. Meanwhile, some systems clarify designers' intentions to support interpretable reasoning [34] and assist reasoning by providing hints and feedback [9]. In this way, human designers and AI can collaboratively enrich and explore design spaces [64].

In this paper, we aim to frame and advance human-AI collaborative DSE through the problem-solution co-evolution model and various design reasoning methods. This research includes three phases. In the formative study involving 12 designers, we adopted the problem-solution co-evolution model as a framework for structuring the design space and applied design reasoning methods to understand the driving force behind DSE. This allowed us to analyze the challenges designers face when navigating the design space with AI and to identify strategies to support collaborative DSE. Then we developed CoExploreDS, a system that formalizes emerging problems and solutions as nodes in a human-AI collaborative design space. CoExploreDS dynamically generates suggestions using four reasoning methods based on the ongoing design process. Finally, we compared CoExploreDS with a baseline system in a between-subjects study involving 32 designers. Our findings show that CoExploreDS enhances design outcomes and fosters creativity in collaborative DSE for two factors: CoExploreDS facilitates more systematic exploration with less effort and appropriately affects designers' reliance on AI and self-confidence in the collaborative process.

In conclusion, this study made three main contributions:

- Modeling the DSE process by proposing a coding method for the design space, analyzing how the design process and designers' attitudes affect collaborative DSE, and proposing three strategies to integrate the problem-solution model and

design reasoning methods into human-AI collaborative DSE in a formative study.
- Developing CoExploreDS, a system that frames and advances the collaborative DSE process by formalizing problems and solutions into nodes and by dynamically generating suggestions based on the ongoing design process using four reasoning methods.
- Demonstrating the utility of CoExploreDS in supporting systematic DSE with lower efforts and positively shaping the human-AI collaboration experience.

# 2 Related Work

## 2.1 Human-AI Collaborative Design

The advancements in large-scale generative models allow AI to increasingly serve as a collaborative design partner for humans [30, 75], significantly altering DSE [64]. Due to their immense size and scale of training data, such models can generate seemingly new content and encompass countless real-world cases [27]. With proficiency in information supplementation and logical reasoning, LLMs like GPT-4 are accelerating collaborative design integration between humans and AI.

Firstly, LLMs help designers integrate scattered knowledge across fields and generate new content, collaboratively enriching the design space. For example, systems such as DiscipLink [79] and GenQuery [63] enable designers to use natural language queries to access broad knowledge bases. Additionally, some systems can automatically suggest conceptual connections for human designers and support evidence-based designs. For example, PopBlends can find conceptual connections between the users' topics and a pop culture domain [71]. Researchers also integrate causal pathway diagrams to support theory-driven design in the domain of human-centered design [81].

Secondly, LLMs have the potential to support reasoning [43, 72]. Reasoning has long been considered a unique human cognitive skill, essential for understanding knowledge and bridging comprehension gaps under constraints [19]. As LLMs gradually exhibit logical reasoning abilities, researchers have been attempting to apply AI to targeted reasoning processes in design. For example, BIDTrainer enables designers to reason more quickly and interactively in bio-inspired design [9]. SituationAdapt can assess the placement of interactive UI elements for contextual UI optimization [46].

Researchers have explored integrating LLMs with classical design methods [10]. For example, systems such as TRIZ-GPT [11] and AutoTRIZ [38] use LLMs to automate and enhance the Theory of Inventive Problem Solving methodology. Guided by the Function-Behavior-Structure model, LLMs can generate more reasonable and creative design concepts [69]. Our paper focuses on the classification and evolution of information within the design space. By framing the overall design space and interpreting designers' movements within DSE, the approach seeks to advance collaborative conceptual ideation processes between designers and AI.

## 2.2 Problem-Solution Co-Evolution in Design

As a descriptive framework, problem-solution co-evolution views DSE as an iterative development process [15, 23]. This theory splits the design space into "problem space" (i.e., the required behavior of the design) and "solution space" (i.e., the potential structural combinations constituting the design). This framework emphasizes that continuous problem formulation and redefinition, paired with solution exploration, drive the evolution and maturation of design [24]. Designers continually propose, evaluate, and reject potential design problems and solutions until a satisfactory "matching" is achieved.

Specifically, four co-evolution episodes were identified: problem evolution (P-P), proposing solutions (P-S), proposing problems (S-P), and solution evolution (S-S). For instance, when designing an electric kettle, the initial problem focused on how to heat water efficiently. As past surveys indicated a desire for precise temperature control for different beverages, the problem evolved into how to provide customizable temperature settings (P-P). To address the new problem, the team proposed integrating a temperature control system with preset options (P-S). After further discussion, a new problem emerged (S-P): how to maintain energy efficiency while offering rapid temperature changes. This led to the solution evolving (S-S) by incorporating advanced insulation and a more efficient heating element. These interactions between problem spaces and solution spaces manifest as cyclical oscillations [23].

Problem-solution co-evolution offers a structured explanation of design space information, essential for clarifying the trajectory of design development [29]. However, the model does not explain the driving forces behind these episodes, especially how designers move from one part to another in the design space. Understanding these forces is essential for maximizing the potential of co-evolution in collaborative DSE.

## 2.3 Reasoning Methods in Design

Design reasoning refers to how designers systematically navigate, process, and organize information under specific constraints within the design space [22]. It involves addressing both objective requirements (e.g., functionality, cost, structural integrity) and subjective requirements (e.g., appearance) [18, 66]. Prior studies have demonstrated that abductive, deductive, and inductive reasoning can explain most step-by-step inferences during the design process [14, 22]. Additionally, analogical reasoning reflects the common cognitive leaps, case-based metaphors, and associative jumps observed in the design process. Designers can use less effort to retrieve and activate past knowledge [7, 61]. These four reasoning methods provide a comprehensive explanation of the design process.

Specifically, deductive and inductive reasoning rely on existing data and do not create entirely new concepts. Deductive reasoning derives solutions from general theories, ensuring adherence to established rules, while inductive reasoning infers concepts from available data within a given model or frame of reference. For example, *To ensure market compliance, the newly designed smart lamp must meet the specific safety standards required for all electronic devices* is an example of deductive reasoning, while *Observing that smartphones with larger screens sell better, a designer might conclude that users generally prefer larger displays* exemplifies inductive reasoning. Besides, abductive reasoning forms new rules or relationships to explain outcomes, making it essential for generating original ideas. For instance, given that a new smartwatch drains its battery quickly, designers might infer that its high-resolution display is the primary power consumer. Additionally, analogical

**Table 1: Examples of design problem-solution co-evolution episodes.**

| Role | Episode Type | Reasoning Methods | Think-Aloud Segments |
|---|---|---|---|
| Designer | Start-Ph | / | "What problems might office workers face?" |
| AI | Ph-Pa | Inductive | "They might be sensitive to price... high expectations for speed, range, safety, and portability. I recommend... long commuting support, portability, easy maintenance, durability." |
| Designer | Pa-Ph | Inductive | "Apart from the needs you mentioned, what other factors might be important for these users? For example, ease of storage and transportation, such as carrying up stairs or using an elevator." |
| AI | Ph-Pa | / | "Stylish design, smart operation... high-efficiency battery, fast charging." |
| Designer | Pa-Ph | Abductive | "Exactly, the point about fast charging is reasonable. Generally, do electric scooters need special charging stations, or can they be charged with regular AC power? Do they require batteries or large capacity storage?" |
| AI | Ph-Sa | Deductive | "They need batteries, but high energy density and compact... Capacity: 9Ah to 30Ah. Battery Management System needed." |

reasoning utilizes similarities to adapt known solutions to new contexts and maps causal structures by comparing the conceptual distance between a source product and the target design problem. *Inspired by the efficient gripping ability of gecko feet, a new type of robotic gripper incorporates micro-suction technology to handle delicate objects safely* is a typical example. By integrating various design reasoning methods, designers constrain or backtrack information in the design space, generating think flows in multiple directions [47, 73].

Due to differences in design experience, preferred reasoning methods, and other cognitive variations, both common and personalized characteristics can be found in designers' reasoning styles [52]. For example, experts employ breadth-first reasoning, while novices adopt depth-first reasoning [16]. Some designers may rely on experiential abstract knowledge and infer patterns from facts, while others tend to rely on case-driven analogies [3]. Therefore, future tools for designers should allow flexible use of design methodologies rather than impose restrictions.

## 3 Formative Study

To understand the challenges designers face when navigating the design space with AI and to identify specific strategies that support DSE, we invited 12 designers to collaborate with GPT-4 on a design task in this formative study. We focused on three key questions: (1) What are the patterns of problem-solution co-evolution when designers and AI collaboratively explore the design space, and how do the patterns influence the DSE process? (2) Are there differences in reasoning methods and design styles of different designers when collaborating with AI, and do these differences impact the DSE process? (3) What are the patterns of designers' attitudes in human-AI collaboration, and do these patterns affect the DSE process?

### 3.1 Participants and Procedure

We recruited 12 participants (D1-D12; age: $M$ = 24.33, $SD$ = 1.34) with backgrounds in product design or industrial design. All the participants had three to five years of experience in design and were familiar with LLM tools for product conceptual design. Participants were required to design a commuting electric scooter within 45 minutes and produce a viable design solution. During the task, they had access to the official OpenAI ChatGPT website platform [1], which utilized the GPT-4 model for text conversations. They were instructed to think aloud during the process [67]. All the participants signed informed consent forms and received monetary compensation after completing the study.

### 3.2 Methods and Metrics

We evaluated the effects of DSE by assessing the quality of final design outcomes. Then, we modeled and analyzed the design process to extract strategies for better DSE. Below, we describe the methods and metrics we used in these analyses.

For the quality of the design outcomes, we utilized expert rating methods. The design outcomes refer to textual solutions summarized by participants at the end of the study. Experts evaluated these outcomes based on two metrics: novelty ($N$) and usefulness ($U$) [36, 62]. For each design outcome, the final scores for $N$ and $U$ were calculated as the mean of the experts' overall scores of the final designs, scaled from 1 to 7. Details of the quality computation are provided in Appendix A.1. We also performed Kendall's W consistency test to ensure consistency among experts. Three experts with over five years of design experience (E1-E3; age: $M$ = 29.33, $SD$ = 4.16) were invited. Each expert was provided with an evaluation form that included descriptions of the metrics to illustrate the rubrics for the rating (Appendix A.2).

For the DSE processes, protocol analysis [21] was employed to examine transcript segments from think-aloud protocols. Our coding themes included two parts. First, we categorized the design process information into four distinct types: problems generated by AI (Pa), solutions generated by AI (Sa), problems generated by human designers (Ph), and solutions generated by human designers (Sh). To further explore the dynamics of problem and solution evolution, we analyzed co-evolution episodes across 16 distinct scenarios, such as human-generated problems to AI-generated solutions (Ph-Sa), AI-generated problems to AI-generated solutions (Pa-Sa), AI-generated problems to human-generated solutions (Pa-Ph), etc. Second, we investigated and coded the usage of four reasoning methods in each episode throughout the design process: abductive, deductive, inductive, and analogical reasoning. Table 1 provides illustrative examples of episodes from selected categories.

To more intuitively analyze patterns and styles in human-AI collaboration, we visualized the collaborative DSE process. We presented the problem-solution co-evolution episodes as nodes arranged in concentric circular layers, where each layer corresponds to an iterative design round, and the layers extend outward to indicate increasing iterations. Circular nodes represent problems, while

[1] https://chatgpt.com/

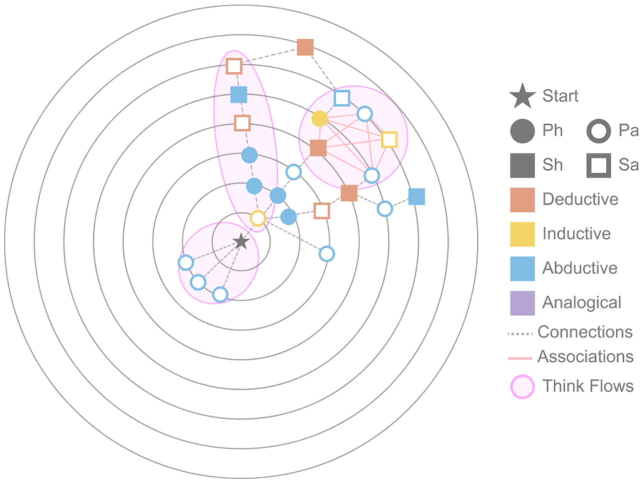


**Figure 2: Visualization method of the process of collaborative DSE. The shapes of nodes represent problems or solutions, and the colors indicate design reasoning methods. Circular layers denote iterative evolution, with nodes added clockwise. Throughout the design process, several think flows are developed.**

square nodes represent solutions. Hollow and solid fills distinguish whether nodes originated from AI or designers, and different node colors denote the associated reasoning methods. A series of related nodes is defined as a "think flow". Figure 2 shows an example of this visualization. To quantitatively evaluate how different think flows influence design outcomes, the quality of the sub-functions generated within each think flow was also rated by the three experts. The experts applied the same $U$ and $N$ metrics used for the overall design outcome.

We conducted statistical analyses to validate the observed patterns and explored relationships between outcome quality and several factors, such as the proportion of problem-solution interactions, the number of reasoning methods employed in a think flow, and the proportion of designer contribution. We used Pearson correlation for normally distributed data and Spearman's rank correlation for non-normal data, reporting correlation coefficients as $r$. Finally, we classified designers' responses to AI suggestions (negative or positive) to gauge attitudes toward AI collaboration.

## 3.3 Findings

Expert ratings demonstrated strong agreement on design outcomes. Specifically, Kendall's coefficients of concordance (Kendall's W) are 0.607 for $N$ and 0.613 for $U$, both statistically significant ($p < .05$). The design outcomes from the 12 designers displayed high novelty ($N = 5.28$, $SD = 1.38$), but there was significant variation in their usefulness ($U = 2.84$, $SD = 1.90$). Examples of the participants' creation outputs and the associated scores are provided in Appendix A.3. In the following sections, we first describe the observed characteristics of human-AI collaboration, and then further discuss possible factors influencing the DSE process.

*3.3.1 Challenges of Problem-Solution Evolution During Human-AI Collaborative Design.* Results showed that the frequency of proposed problems and solutions by human-AI teams is comparable to those of human-human teams. Specifically, designers and AI collectively mentioned design problems 250 times (59.1%, including Pa and Ph) and proposed solutions 173 times (40.9%, including Sa and Sh), closely mirroring the 55.02% and 44.98% reported for human-human teams in prior research [73]. Thus, the similarity in using the problem-solution co-evolution model between AI and designers suggests that this theory can be transferred to systems for collaborative DSE as a descriptive framework for the design space.

**Compromise of the DSE process due to infrequent design space interactions**. Figure 3(A) shows that limited interactions between the problem space and the solution space may hinder DSE, as evidenced by lower-quality design outcomes. Think flows with lower-quality outcomes typically exhibit only a one-way transition from the problem space to the solution space. In contrast, think flows with higher-quality outcomes often involve four or more problem-solution interactions. Statistical analyses further support this observation: the frequency of problem-solution interactions within a think flow correlated positively with outcome quality, both for $N$ ($r = 0.833$, $p < .01$) and $U$ ($r = 0.801$, $p < .01$).

*3.3.2 Challenges of Reasoning Methods Usage.* The results manifested that designers and AI tend to employ different reasoning methods. According to the analysis of experimental data, humans and AI were engaged in reasoning 294 times (See details in Appendix A.4). Among them, abductive reasoning was prevalent in the problem space evolution, with 84 instances recorded (designers:44, AI:40). Inductive reasoning also featured prominently, with 51 instances (designers:18, AI:33). In the solution space, deductive (51 instances, designers:10, AI:41) and abductive reasoning (49 instances, designers:14, AI:35) were more frequently observed. Analogical reasoning was used the least (only 14 instances, all initiated by designers), but often led to significant breakthroughs. Next, we introduce two potential challenges arising in human-AI collaborative reasoning within DSE.

**Negative impact of path dependency of reasoning on the DSE process**. Designers often employ consistent reasoning methods within each design iteration and adhere to a specific logical sequence across various solutions ("Preferred logic" in Figure 3(B)). This path dependency may negatively affect the DSE process, as proved by lower-quality outcomes. Figure 3(A) supports that integrating diverse reasoning methods within the same iteration and think flow is associated with the enhanced DSE process. Additionally, we counted the number of think flows per participant and calculated the proportion of flows that included more than three types of reasoning methods. This proportion showed a significant correlation with $N$ ($r = 0.682$, $p < .05$) and a marginally significant correlation with $U$ ($r = 0.567$, $p = .054$). Based on these findings, we suggest avoiding path dependency of reasoning methods to enhance DSE.

**Influence of think flow types on the DSE process**. To describe design styles in reasoning, we categorized three types of think flow by shape: linear, divergent, and associative (Figure 3(C)). Designers often combine different think flows in DSE (Figure 3(B)),

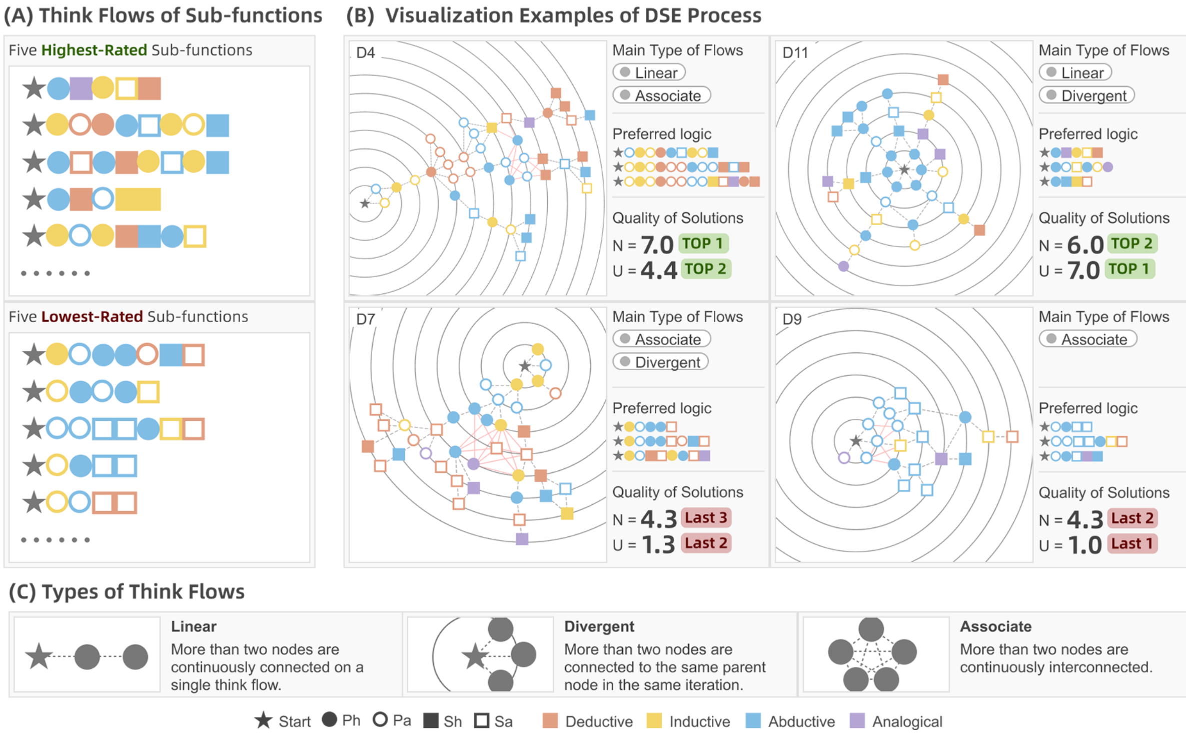


**Figure 3: Visualization of the process of collaborative DSE. (A) displays the think flows from which the five highest-rated and five lowest-rated sub-functions are derived. (B) illustrates some examples, including the main types of think flows that constitute the node map, designers' preferences for reasoning methods, and the novelty and usefulness of design outcomes. (C) shows three types of think flows, including linear, divergent, and associative.**

such as linear with divergent (e.g., D11), linear with associative (e.g., D4), or divergent with associative (e.g., D7 and D9). However, design processes dominated by divergent and associative think flows tend to hinder comprehensive DSE, associating with lower-quality overall outcomes ($N = 4.33$, $U = 1.92$) compared to the average of the other two types (linear with divergent: $N = 6.00$, $U = 4.32$; linear with associative: $N = 6.07$, $U = 3.31$). This may be because these flows result in a loose design process, making it hard for designers to converge on specific solutions. Therefore, it is advisable to avoid consistently using divergent and associative reasoning, instead favoring combinations of linear with divergent or linear with associative to facilitate a more structured DSE.

*3.3.3 Impact of Designers' Collaborative Attitudes.* In Figure 3(A), we observed that human contributions in the five highest-rated sub-functions (24 designer contributions: 9 AI contributions) appear to be more frequent than those in the lowest-rated outcomes (8 designer contributions: 19 AI contributions). However, the proportion of contributions by human designers did not show a significant correlation with $N$ ($r = 0.372$, $p > .05$) or $U$ ($r = 0.177$, $p > .05$). Based on the interviews and experimental observations, this difference may stem from the dynamic relationship between collaborative attitudes in human-AI design, particularly designers' self-confidence and their reliance on AI.

**Two inappropriate attitudes in human-AI collaboration**. Designers who consistently oppose AI may also hinder sufficient human-AI interactions in the design space. For example, D7, who was overly confident and had biases against AI's suggestions, frequently expressed doubts with statements like "*I think you're wrong...*" and "*I don't agree...*". This was reflected by an incomplete DSE, as evidenced by the second and third lowest scores ($N = 4.33$, $U = 1.26$) in Figure 3(B). Meanwhile, designers who overly rely on AI suggestions to complete designs (e.g., D9) tend to confine themselves to a narrower design space with an illusion of success [77], which was also associated with the lowest-quality outcomes ($N = 4.33$, $U = 1.00$).

**A dynamic relationship between collaborative attitudes and the DSE process**. In AI-assisted DSE, collaborative attitudes "*dynamically change in response to AI's real-time feedback and design proposals*" (D1 and D5). These changes depend on the quality of AI-generated content, the transparency of the creation process, and the alignment with designers' intentions. A designer's self-confidence and reliance on AI typically reflect a self-assessment of the outcome's quality. This self-assessment affects the designer's exploration of the design space, thereby impacting the ultimate quality of the solutions. To ensure effective collaboration and prevent over-reliance or overconfidence, AI must provide appropriate guidance to enhance the DSE process.

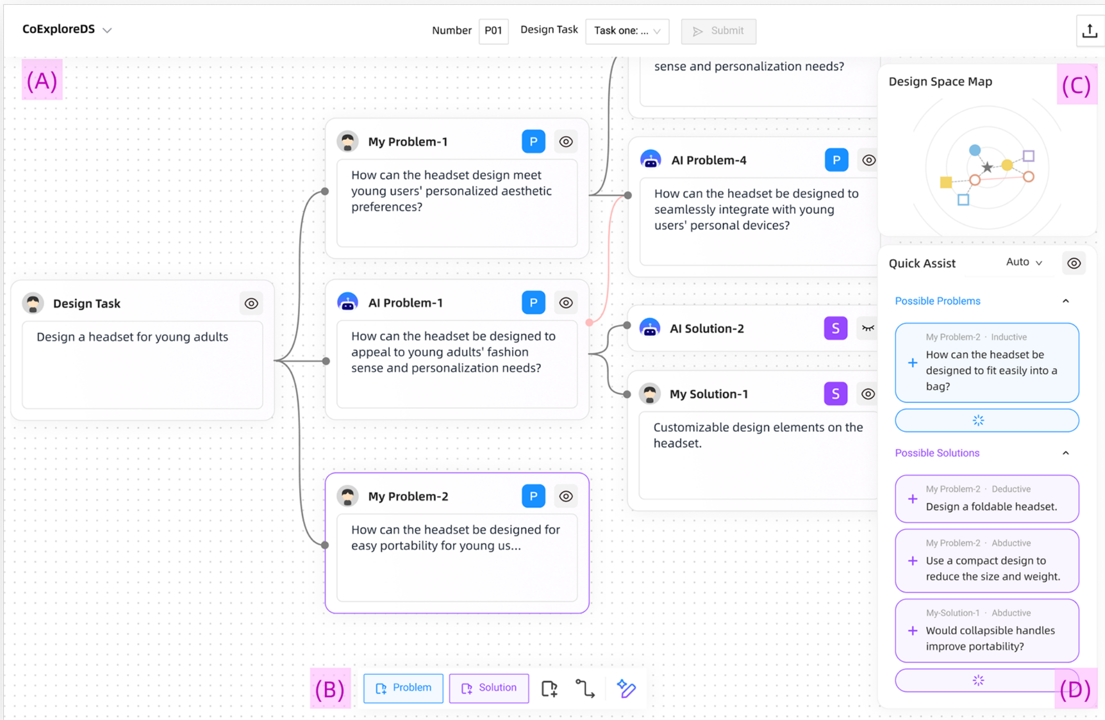


**Figure 4: Overview of CoExploreDS. (A) Main Canvas: Designers can add, move, hide, and connect nodes on a mindmap-style canvas. (B) Bottom Toolbar: Provides tools for node operations and text generation via AI initiated by designers. (C) Real-Time Design Space Map Display: Shows how AI understands the designer's process. (D) Quick Assist: For the selected node, it actively offers AI-generated suggestions (e.g., "Possible Problems" and "Possible Solutions") based on the design space map and generation strategies. Clicking a suggestion adds the corresponding node to the main canvas.**

## 3.4 Design Goals

In conclusion, we observed similar problem and solution frequencies between human-AI and human-human teams, indicating the reasonableness of modeling human-AI collaborative DSE with the problem-solution co-evolution model. However, for better DSE, we also identified the necessity of guiding AI systematically in advancing the co-evolution process and prompting designers to adopt appropriate collaborative attitudes. Based on previous findings, we defined three design goals (DG) to develop a system that enables collaborative DSE between designers and AI.

- **DG1: Increase the frequency of interactions between problem space and solution space.** The system should facilitate more frequent exchanges between the problem space and the solution space to improve the iterative design process, ensuring that solutions are continuously refined and aligned with evolving problem definitions.
- **DG2: Utilize various reasoning methods within the same iteration or think flow.** The system should encourage designers to integrate multiple reasoning methods to comprehensively explore the design space. In addition, it should guide designers to avoid combining divergent thinking with associative thinking flows, which may lead to an overly loose exploration process that hinders the ability to converge on definitive solutions.
- **DG3: Provide an appropriate mode of AI intervention.** The system is supposed to guide designers to adopt suitable attitudes toward AI, such as self-confidence and reliance appropriateness, thereby avoiding limited DSE caused by biases or illusions.

## 4 CoExploreDS

We developed CoExploreDS, a system designed for human-AI collaborative DSE. CoExploreDS leverages the problem-solution co-evolution model to structurally frame the design space. This system also incorporates deductive, inductive, abductive, and analogical reasoning methods to proactively generate problem or solution suggestions. To differentiate and display problems and solutions, CoExploreDS uses uniquely styled card-like nodes to clarify their classification. Designers can actively request AI to generate a problem node or solution node for a specific node (passive generation) or apply suggestions actively provided by AI to the canvas (proactive suggestions). The overview of CoExploreDS is shown in Figure 4, and it integrates four panels: the main canvas (Figure 4(A)), the bottom toolbar (Figure 4(B)), a real-time design space visualization map display (Figure 4(C)), and a Quick Assist panel (Figure 4(D)).

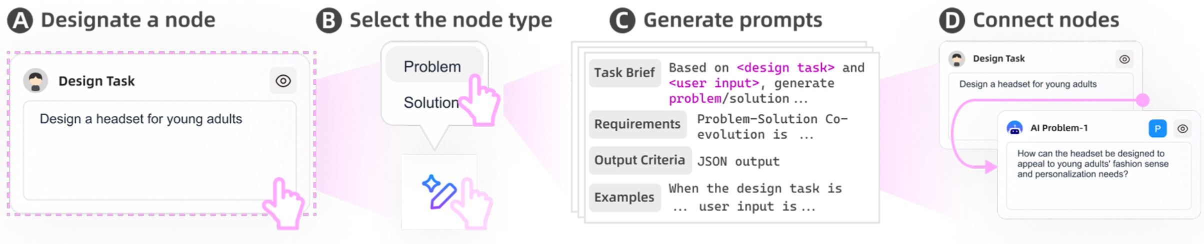


**Figure 5: Pipeline for passive AI node generation instructed by designers. When designers instruct AI to generate new content, they need to (A) select a node as input and (B) decide whether the node type to be generated is a problem or a solution. CoExploreDS's backend then (C) transforms the user input into structured prompts, including task briefs, requirements, output criteria, and examples. (D) The newly added AI node is placed on the main canvas and connected to the source node.**

## 4.1 Node-Based Collaborative Design Canvas

CoExploreDS employs a mind map format to organize nodes on the main canvas (Figure 4(A)), distinguishing between problem nodes and solution nodes. The mind map format is used on the main canvas because it effectively organizes design information [26]. Designers can drag and zoom the main canvas. To manipulate nodes, five features are integrated into the bottom toolbar (Figure 4(B)): creating a new problem node, creating a new solution node, creating a blank node, connecting two nodes, and instructing AI to generate content. When designers intend to add new nodes to the canvas, they are required to first designate a node as the parent node. Once a new node is added, it automatically links to its parent. To connect two nodes, designers select a source node and then a target node, creating a directed link. This linking process aids in clarifying designers' thought processes and helps AI understand the relationships between the nodes envisioned by the designers.

Figure 5 illustrates the pipeline for passive AI node generation. Designers can instruct AI to generate either a problem node or a solution node based on a specific node. When designers instruct AI to generate new content, they must select a node as input (Figure 5(A)) and decide whether the node type to be generated is a problem or a solution (Figure 5(B)). CoExploreDS's backend transforms the user input into structured prompts. Based on prior studies [60], the prompts for content generation include task briefs, requirements, output criteria, and examples (Figure 5(C)). The newly added AI node is placed on the main canvas and connected to the source node (Figure 5(D)).

## 4.2 Design Space Structured Visualization Maps

CoExploreDS visualizes the collaborative DSE process in the design space structured using layered maps (Figure 4(C)). The nodes' forms and meanings in these maps follow the approach described in Section 3.3.2. Node shapes indicate whether a node is a Problem or a Solution, while the fill status (solid or hollow) denotes whether it was manually created by the designer or generated by the LLM. Connections between nodes are derived directly from the main canvas, and node colors indicate the reasoning methods used, as determined by the LLM. Each modification to the main canvas updates the design space map accordingly. These maps are generated from JSON data that contains all problem and solution nodes and their relationships. We developed a custom algorithm to establish adjacency relations and assign levels to each node, producing the final layered representation of the design space. Pseudocode for this algorithm is provided in Appendix B.1.

## 4.3 Proactive Suggestions Based on Design Space Maps

When designers click on a node, the system will proactively suggest "Possible Problems" and "Possible Solutions" using appropriate reasoning methods based on the designer's current think flow and design content. Specifically, the LLM receives JSON data describing all problem and solution nodes and their relationships on the canvas. Then, the LLM generates suggestions displayed in the Quick Assist panel. The pipeline for AI to proactively generate suggestions is illustrated in Figure 6 and the pseudocode is provided in Appendix B.2.

The system identifies the node's think flow type—linear, divergent, or associative—and generates new nodes accordingly (DG2). For linear or divergent flows, two new nodes are generated at the same level and sub-level to help designers explore comprehensively. For associative flows, four nodes are generated at the sub-level to prevent a fragmented thinking process. Based on these boolean values and basic node information, potential positions are generated, resulting in a list that contains information on which node the new node will connect to.

Additionally, when the selected node is a problem node, more "Possible Solutions" are actively provided to promote interactions between the problem space and the solution space; the reverse applies when editing a solution node (DG1). Then based on the previous node connected to the one currently being edited by the designer, the reasoning methods used for suggestions are prioritized as follows: for generating "Possible Problems", the priority is abductive > inductive > deductive; for generating "Possible Solutions", the priority is deductive > abductive > inductive. An analogical reasoning suggestion is also always generated, regardless of the type (DG2). To avoid fostering an inappropriate collaborative attitude, CoExploreDS enhances collaboration transparency through the design space map and strengthens the designer's agency with proactive and passive AI assistance (DG3).

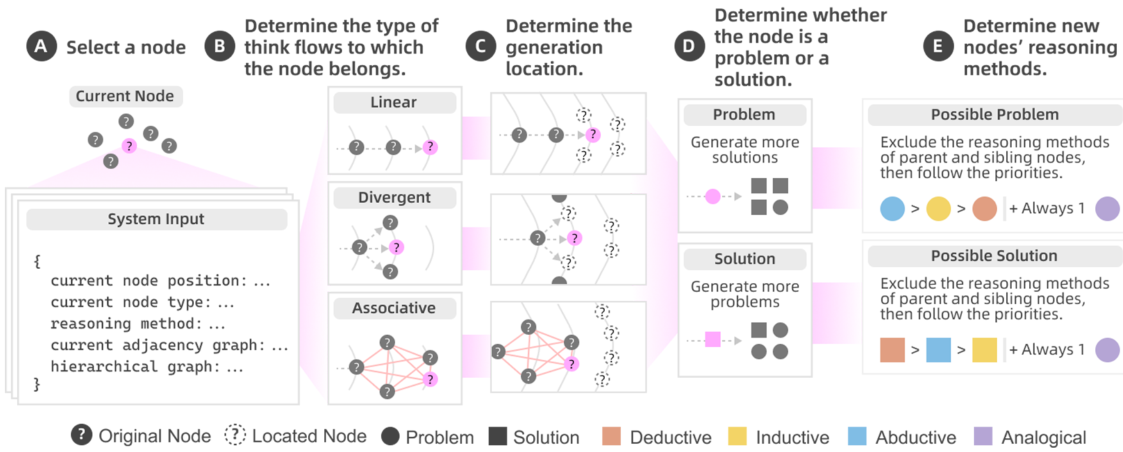


**Figure 6: Pipeline for proactive AI suggestions in the Quick Assist panel. When a designer (A) edits a node card on the main canvas, CoExploreDS receives input. The system then (B) determines the type of think flows to which the node belongs, (C) determines the generation location, (D) assesses whether the node is a problem or a solution, and (E) identifies the new nodes' reasoning methods before generating outputs.**

Finally, CoExploreDS adds the generated suggestions to the list of all possible node generation information for return. All the generated suggestions will be shown in the Quick Assist panel. Each suggestion card indicates the position, reasoning method, problem or solution type, and specific content. Suggestions are added to the canvas only when the designer clicks on them.

## 4.4 Implementation

CoExploreDS is a web-based platform developed with a ReactJS[2] front-end and a Flask[3] back-end. It utilizes Ant Design[4] for styling. Text generation is managed by GPT-4[5] through the OpenAI API, chosen for its nuanced understanding of user intents and capability in handling complex tasks. Technical pre-testing has demonstrated the GPT model's high accuracy in identifying reasoning methods and understanding the problem-solution co-evolution.

# 5 User Study

To validate the effectiveness of CoExploreDS in supporting human-AI collaborative DSE, we conducted a between-subjects experiment with 32 participants against a baseline system. Specifically, we raised the following research questions:

- **RQ1:** Can CoExploreDS systematically support DSE with lower efforts of human-AI collaboration?
- **RQ2:** Can CoExploreDS appropriately affect designers' self-confidence and reliance on AI during human-AI collaborative DSE?

[2]https://react.dev/
[3]https://flask.palletsprojects.com/
[4]https://ant.design/
[5]https://openai.com/gpt-4

## 5.1 Participants and Procedure

We recruited 32 participants (P1-P32; age: $M = 23.91$, $SD = 1.23$) via social media, each with over three years of product design experience and no prior involvement in the formative study. All the participants were evenly divided into two groups, ensuring comparability regarding age, design experience, and gender. Demographics of participants and their groupings are provided in Appendix C.1. The baseline system used was a simplified version of CoExploreDS, retaining only the mindmap function and basic AI generation capabilities, with the "co-evolution" and "reasoning patterns" functionalities removed (Appendix C.2). In the baseline system, the text content of a user-selected node was used as a prompt to retrieve GPT-generated responses via an API, which were then displayed on the canvas in a new node. The 60-minute procedure is outlined below:

**Introduction (10min)**. Participants signed an informed consent form and were introduced to the study context. Subsequently, participants watched the tutorial video for either CoExploreDS or the baseline system according to their assigned group. P1 to P16 used CoExploreDS, while P17 to P32 used the baseline system.

**Ideation (30min)**. Participants in the two groups completed a 30-minute ideation task using either CoExploreDS or the baseline system. To eliminate potential task-related biases, half of the participants worked on Design Task A, while the other half worked on Design Task B. Design Task A was to design a delivery drone for urban and suburban areas. Design Task B was to design a household food processor capable of quickly performing tasks such as mixing and cooking ingredients.

**Post-experiment Survey (20min)**. Participants were required to fill out a questionnaire and undergo a semi-structured interview at the end (Appendix C.3). All the participants received monetary compensation of RMB 50 after completing the study.

## 5.2 Methods and Metrics

To evaluate the final effects of the DSE process, we assessed the quality of design outcomes and discussed CoExploreDS's role in enhancing creativity using the Creativity Support Index (CSI) scale [12]. Additionally, the same three design experts mentioned in Section 3.1 were enlisted to rate the design outcomes. The rating form for experts was also the same (Appendix A.2). The overall scores of design outcomes (i.e., $N$ and $U$) were measured through self-assessments and expert evaluations to ensure reliability, consistent with the methods described in Section 3.2. The same DSE process visualization was also used in the user study.

Then we employed targeted metrics for each RQ. For RQ1, the System Usability Scale (SUS) [5] evaluated information usability during the design process, and the weighted overall score of the NASA Task Load Index (NASA-TLX) [33] evaluated designers' effort expended in the collaboration. For RQ2, a seven-point scale measured designers' confidence and reliance (Appendix C.4).

For quantitative data comparing CoExploreDS and the baseline system, we initially assessed distribution normality using the Shapiro-Wilk test. For normally distributed data, independent-sample t-tests were conducted to compare group means, with the assumption of equal variances validated using the F-test. For non-normal datasets, we applied the Mann-Whitney U test. Qualitative data from interviews were coded into themes to explain quantitative data further.

# 6 Result

Results indicated that CoExploreDS facilitated human-AI collaborative DSE, as demonstrated by significantly improving design outcomes and enhancing creativity. Specifically, CoExploreDS enabled designers to systematically explore the design space using the problem-solution co-evolution model and design reasoning methods, thereby significantly reducing the effort required in human-AI collaboration. Meanwhile, our system also affected designers' reliance on AI and their self-confidence in human-AI collaboration to support collaborative DSE. An example of the participants' design processes with CoExploreDS is shown in Appendix C.5.

## 6.1 Generating More Creative Design Outcomes in DSE

*6.1.1 Quantitative Evidence for Enhanced Creativity and Quality in DSE.* In terms of both the quality of its outcomes and the processes of DSE, CoExploreDS outperformed the baseline system. The significantly higher quality of design outcomes generated by participants using CoExploreDS supported that our system fully unleashed the potential of human-AI collaboration in DSE (Table 2). Expert evaluations demonstrate consistency, with Kendall's W values for CoExploreDS at 0.60 ($N$, $p < .05$) and 0.57 ($U$, $p < .05$), and for the baseline system at 0.757 ($N$, $p < .01$) and 0.796 ($U$, $p < .01$). For metric $N$, expert ratings revealed that CoExploreDS achieved a higher mean score of 5.00 compared to the baseline's 4.68, with a p-value of .033. The participants' self-reported $N$ was also higher in the CoExploreDS system than in the baseline system (5.31 vs. 4.69), with a p-value of .018. Similarly, for metric $U$, CoExploreDS showed higher mean scores in self-assessment (2.98 vs. 2.69; $p < .05$) and expert ratings (3.19 vs. 2.78; $p < .05$).

**Table 2: Comparison of self-assessment and expert ratings of design quality between CoExploreDS and the baseline systems (*: $p < .05$).**

| | | Self-Assessment | | Expert Rating | |
|---|---|---|---|---|---|
| | | CoExploreDS | Baseline | CoExploreDS | Baseline |
| $N$ | $M$ | 5.31 | 4.69 | 5.00 | 4.68 |
| | $SD$ | 0.87 | 0.48 | 1.20 | 1.91 |
| | $p$ | .018(*) | | .033(*) | |
| $U$ | $M$ | 2.98 | 2.69 | 3.19 | 2.78 |
| | $SD$ | 1.27 | 0.71 | 1.03 | 0.38 |
| | $p$ | .030(*) | | .029(*) | |

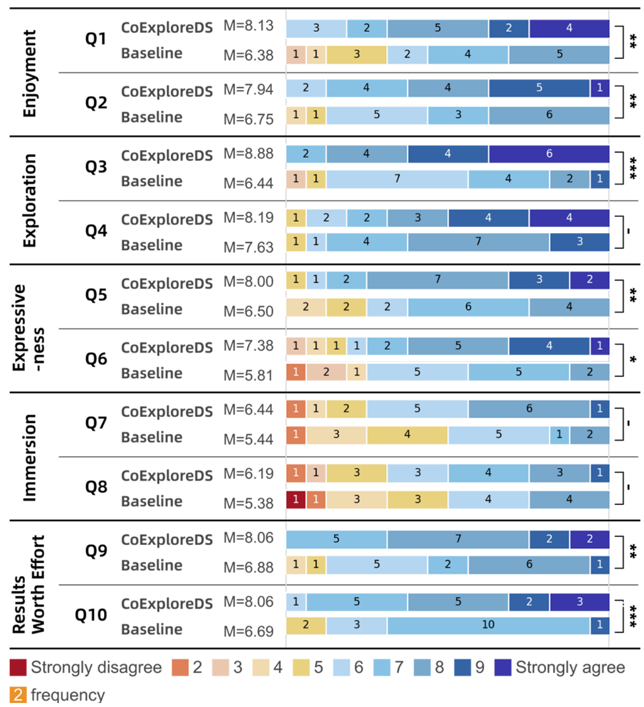


**Figure 7: Results of the CSI questionaires (-: $p > .05$, *: $p < .05$, **: $p < .01$, ***: $p < .001$).**

For the design process, the CSI results indicated that CoExploreDS ($M = 74.34$, $SD = 9.00$) significantly enhanced creativity compared to the baseline system ($M = 66.63$, $SD = 9.94$), with a p-value of .028. This demonstrates CoExploreDS's impact on fostering creativity over the baseline system (Figure 7). CoExploreDS supported more frequent co-evolution episodes (39.63 instances per participant) than the baseline system (34.94 instances per participant). Although the baseline generated fewer nodes, the difference between the two groups was not significant ($p > .05$), suggesting that the impact of the QuickAssist and design maps extended beyond merely accelerating the design process. Additionally, all participants demonstrated at least three different combinations of design reasoning methods when using CoExploreDS. In contrast, their reasoning predominantly adhered to similar ideation logic in the baseline system.

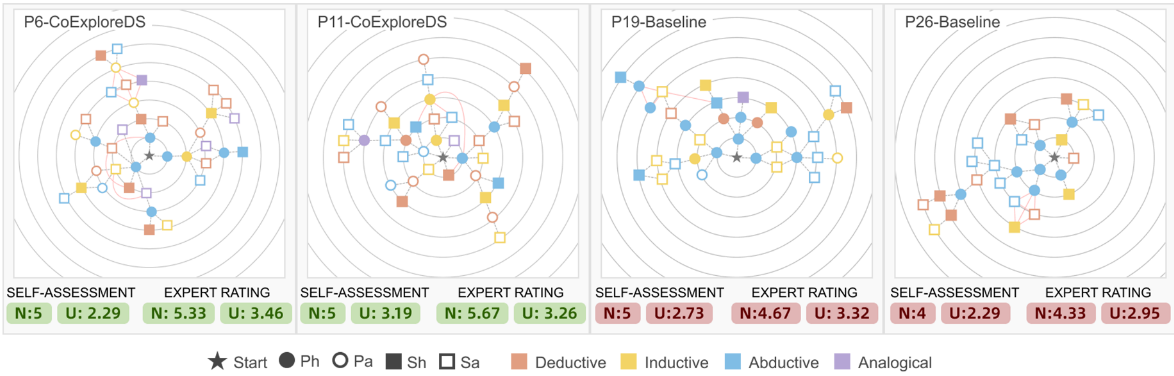


**Figure 8: Design quality and design space maps of several participants when using CoExploreDS and the baseline system.**

*6.1.2 Underlying Reasons for Improved Creativity and Quality.* Interview results showed that participants generally praised CoExploreDS for its suggestions' quality and proactive interactions in supporting DSE. P13 noted, "*Thanks to the application of various design reasoning methods, the content generated by AI in CoExploreDS is more diverse.*" P7 also added, "*New content continuously emerges during the iterative process between problem space and solution space in CoExploreDS.*" Meanwhile, P15 mentioned, "*An AI that proactively interacts at key moments can greatly boost creativity compared to working alone.*" In contrast, participants using the baseline system expressed concerns regarding the AI's limitations in relevance and the alignment of generated content with design needs. For example, P24 and P27 stated that AI primarily "*performed knowledge relocation without considering the feasibility of the solutions.*" Additionally, four participants observed that "*the content provided by AI lacked depth and was often impractical.*" P24 further remarked that "*the AI's suggestions were overly focused on technical aspects, leading to a deviation in my design process.*"

*6.1.3 User Cases for Better DSE.* Considering a fair and representative comparison, participants with scores of $N$ and $U$ close to the mean of the group were selected as typical cases to examine the effects of CoExploreDS and the baseline system on DSE (Figure 8). Using CoExploreDS to complete the DSE, P6 (Expert rating: $N = 5.33$, $U = 3.46$) and P11 (Expert rating: $N = 5.67$, $U = 3.26$) employed various reasoning methods within the same iteration or think flow and had frequent interactions between problem spaces and solution spaces. Furthermore, the four reasoning methods—abductive, inductive, deductive, and analogical reasoning—appeared in a more random order during their DSE process. For example, P6 considered why drones have not yet been widely adopted and identified the question of "*how to control costs and improve delivery efficiency*" (abductive reasoning). AI proposed a solution centered on "*finding a balance among material costs, body weight, rotor size, and battery motor configuration*" (deductive reasoning). After multiple iterations, P6 deduced a sub-function of "*providing modular cargo boxes and hooks for different types of goods*" to meet various needs and reduce operational costs. However, P19 (Expert rating: $N = 4.67$, $U = 3.32$) and P26 (Expert rating: $N = 4.33$, $U = 2.95$) using the baseline system demonstrated more one-way transitions from problems to solutions, and the think flows on the maps usually consisted of fixed sequences of reasoning methods. P26 commented that "*AI directly provides solutions rather than guiding the designer's thinking*" and that "*sometimes the AI's contributions lack thoughtfulness or feasibility.*" All these findings further validated the strategies proposed for better collaborative DSE in our formative study.

## 6.2 Supporting Systematic DSE with Lowered Collaboration Effort

*6.2.1 Quantitative Evidence for Lowered Collaboration Effort.* According to the SUS standard, both CoExploreDS ($M = 81.88$, $SD = 11.49$) and the baseline system ($M = 73.90$, $SD = 8.19$) achieved usability scores above the acceptance threshold of 70. CoExploreDS scored significantly higher in usability than the baseline system ($p < .05$). As suggested by NASA-TLX scores, CoExploreDS ($M = 36.54$, $SD = 10.99$) also significantly outperformed the baseline system ($M = 54.85$, $SD = 13.12$) for workload ($p < .001$). This result demonstrates that designers expended considerably less effort in human-AI collaborative DSE using CoExploreDS.

*6.2.2 Underlying Reasons for Enhanced Effect.* We concluded three reasons for the lowered collaboration effort. First, CoExploreDS "*provides information closely aligned with the design intent*" (P1, P5, P6). Second, the design space map in CoExploreDS "*helps designers identify and address potential gaps in their thinking, reducing the cognitive load needed*" (P8, P11). "*When unsure what to do next, I open the map and turn to expand the shorter think flows*" (P2, P8, P14). Third, CoExploreDS is more responsive, "*aligning better with the typical iterative design process that moves between problems and solutions*" (P7, P16), and "*It actively suggested more relevant problems related to my proposed solutions as I developed them*" (P2). In contrast, the baseline system required designers to actively engage in a back-and-forth with AI to advance the design process. P19 remarked that the system "*could not effectively follow the designer's thought process, particularly when the thinking was broad or highly non-linear*".

*6.2.3 User Cases for Systematic and Effective DSE.* We compared design space maps between two systems and analyzed participants with NASA-TLX scores near the group mean to demonstrate CoExploreDS's support for collaborative DSE processes. Figure 9 displays NASA-TLX scores and design space maps for participants using CoExploreDS and the baseline system. In the design space maps

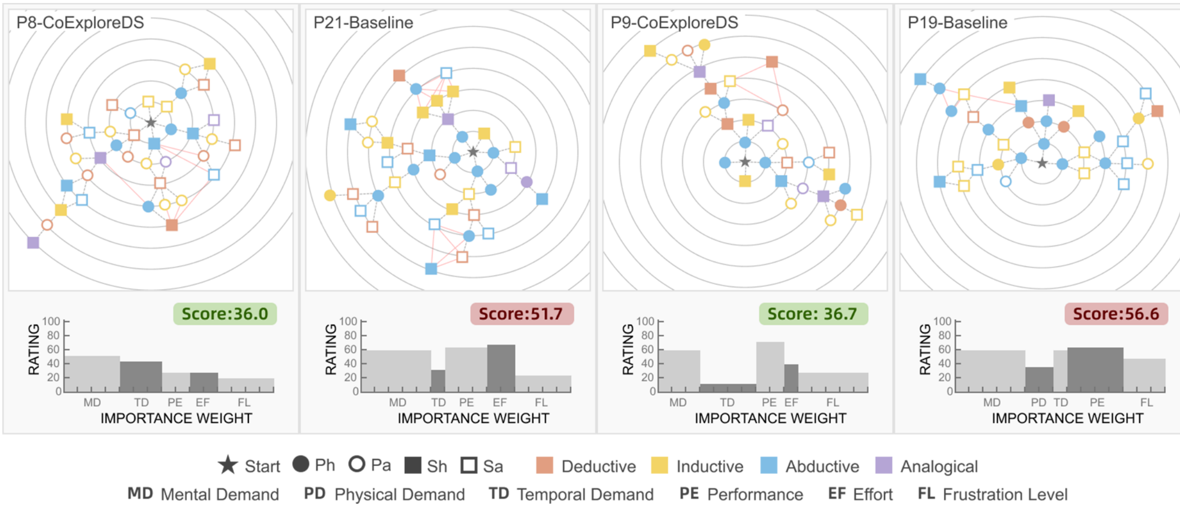


**Figure 9: Example results of NASA-TLX scores and design space maps.**

of P8 and P21, the number of iterations was similar, yet P8 with CoExploreDS experienced a lighter workload (36.0) compared to P21 with the baseline system (51.7). Both participants were highly sensitive to Mental Demand, weighing 4 and 5. P8 using CoExploreDS noted that "*AI acts like a vast knowledge base, uncovering unexpected ideas. It offers expertise on drones and draws from design experiences in other fields. This reduces my workload as I only need to filter, assess, and iterate on these suggestions.*" In contrast, P21 using the baseline system noted, "*AI provides little substantive content and often repeats the same information, requiring me to exert more effort.*" In the design maps of P9 and P19, the differences in NASA-TLX scores likely stemmed more from Performance and Frustration Levels. P9 mentioned, "*Using CoExploreDS made the design process smoother. When considering a problem, CoExploreDS anticipated my need and generated more solutions, giving me more time to iterate and improve my designs.*" Prior research suggests design ideation relies on long-term memory retrieval [57], and that capturing and curating design information is crucial for process improvement [40]. By framing and advancing the DSE process, CoExploreDS enables designers to retrieve and integrate design information with fewer cognitive resources, allowing for greater focus on creation.

## 6.3 Affecting Designers' Self-Confidence and Their Reliance on AI Appropriately

*6.3.1 Quantitative Evidence for Attitude Appropriateness.* As shown in Figure 10(A), participants using CoExploreDS reported significantly higher self-confidence in their design outcomes ($M$ = 5.44, $SD$ = 0.73) compared to those using the baseline system ($M$ = 4.75, $SD$ = 0.93), with a p-value of .038. Additionally, we observed a higher level of reliance on AI ($M$ = 5.63, $SD$ = 1.31) in comparison to the baseline system ($M$ = 4.49, $SD$ = 0.93), with a p-value of .023.

*6.3.2 Relationship Between Attitudes and the DSE Process.* To investigate how human reliance on AI and self-confidence influence DSE within CoExploreDS, we generated scatter plots. As described in Section 5.2, we chose the quality of design outcomes to quantify the effectiveness of the DSE process. Figure 10(B) illustrates the relationships between design quality and reliance on AI, and Figure 10(C) depicts the relationship between design quality and self-confidence. Design quality is quantified as the sum of $N$ and $U$ rated by experts. The scatter plots are divided into four quadrants based on the mean values of CoExploreDS and the baseline system.

In Figure 10(B), participants using CoExploreDS demonstrated greater reliance on AI, yet their design quality was higher, with most falling within the "High Quality" Quadrants B1 and B2. In contrast, 11 participants using the baseline system clustered in Quadrants B3 and B4, showing lower design quality. This outcome underscores the effectiveness of CoExploreDS in unlocking the creative potential of human-AI collaboration. Participants discussed reasons for this. When ideating with the baseline system, P18, P30, and P31 noted that the AI "*lacked proactive guidance, requiring designers to generate ideas first before instructing the AI. By that point, the designers had already formulated the ideas themselves, leaving the AI unable to contribute spark thoughts.*" In contrast, with CoExploreDS, participants "*frequently utilized AI and effectively guided it to generate more valuable suggestions*" (P1, P3, P15).

Ideally, if the quality and process of AI-generated content meet the designers' expectations, their self-confidence should correspond to the actual quality of the design, thus promoting a beneficial DSE environment that yields higher-quality solutions. In Figure 10(C), Quadrant C1 represents the "High Confidence-High Quality" scenario, the optimal state for human-AI collaboration. This quadrant includes five participants using CoExploreDS, compared to only one participant using the baseline system. Four participants using CoExploreDS fall into Quadrant C2, the "Low Confidence-High Quality" area, yet their confidence levels remain relatively high (around 5 points). Conversely, seven participants using the baseline system predominantly cluster in Quadrant C4, indicating generally lower confidence and design quality.

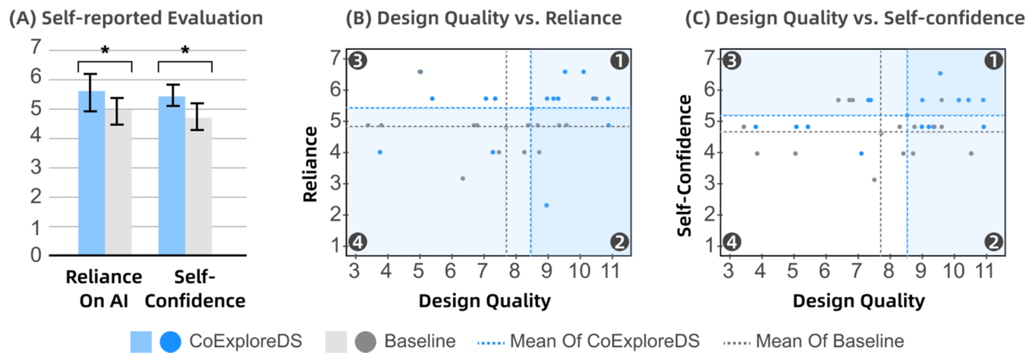


**Figure 10: Self-reported evaluation results for CoExploreDS and the baseline system. (A) Scores of reliance on AI and self-confidence with 95% confidence intervals (*: $p < .05$). (B) Scatter plot of design quality versus reliance for both systems. (C) Scatter plot of design quality versus self-confidence for both systems. In (B) and (C), design quality is represented by the sum of $N$ and $U$ as rated by experts.**

## 7 Discussion

We propose CoExploreDS to support human-AI collaborative DSE by framing and advancing human-AI collaborative DSE through the problem-solution co-evolution model and design reasoning methods. Based on our findings, we suggest some design implications for future human-AI collaborative creativity support systems.

### 7.1 Freeing Designers from Consciously Following Design Methodologies with AI Assistance

Designers often engage in extensive theoretical study and practical training to master navigation through ambiguous and dynamic design spaces [66]. While in practice, designers often depend more on their experience and intuition than on strict adherence to valuable methodologies in complex and varied practices [54]. In our studies, all participants acknowledged the principles and utility of design reasoning methods, yet nearly all lacked awareness of their utilized methods, with only two recalling their methodologies accurately in the post-survey. This suggests that methodologies are often implicitly integrated into the designer's experiential intuition in their daily work [2].

While such synthesis can enhance problem-solving efficiency by allowing designers to act fluidly, it may also lead to overly "opportunistic" ideation, particularly in complex design practices where uncertainty is prominent [16]. In short-term thinking, designers may inadvertently overlook certain considerations, requiring structured methodologies to supplement and guide the process. Therefore, design involves an adaptive combination of both structured and opportunistic processing, where the dynamic balance between the two enables designers to navigate uncertainty and complexity effectively [68].

To accommodate the characteristics of design, our work proposes a flexible solution: AI can subtly guide designers in adhering to methodologies rather than "*directly generating complete, logically unverifiable solutions*" (P6). This approach liberates cognitive resources, enabling designers to focus on exploring solutions without the burden of consciously following methodologies. Additionally, designers avoid the challenge of issuing precise instructions for AI to understand their thought processes [76]. This solution can also alleviate widespread concerns about AI-based creativity support systems to rapidly produce large-scale, low-cost design solutions that may encroach on designers' space [1, 51]. Our study highlights two benefits of this guidance. Firstly, it promotes personalization in the design process, as evidenced by the distinct stylistic characteristics in participants' design space maps. Secondly, it ensures higher-quality outcomes in DSE.

### 7.2 Transferring Systematic Thinking to Other Human-AI Collaboration Innovation Tasks

The results showed that CoExploreDS enhanced DSE through AI suggestions based on design space maps, thereby validating key design goals proposed in our formative study. Guided by problem-solution co-evolution and design reasoning methods, the systematic thinking offered by design space maps facilitated a comprehensive understanding of the collaborative design process, enabling AI to "*generate more contextually relevant and innovative suggestions*" (P1, P5, P6). This systematic thinking also "*promoted creativity by identifying connections and insights that may be overlooked with a more localized approach*" (P11). When designers unconsciously become overly immersed in either breadth-first or depth-first search, the proactive suggestions at different levels of the think flows can encourage them to adopt a more systematic perspective.

Systematic thinking is essential for fostering creativity within human-AI collaborative innovation tasks. These tasks often demand high levels of innovation within multiple constraints [4, 66]. Thus merely recording and reusing previous designs falls short for AI, as this approach inadequately addresses the complexities of current design challenges [70]. As a potential solution, AI equipped with systematic thinking could flexibly and effectively integrate breadth-first and depth-first design strategies in a context-sensitive way, aligning with the better design thinking methods recognized by previous researchers [2]. This approach ensures the co-development of problem frames while facilitating a balanced search for solution alternatives. Similarly, researchers have proposed other solutions, such as improving DSE to discover unexpected regions and assist workflows [17, 82], or providing comprehensive frameworks spanning macro to micro levels [64]. These methods have already been applied to other high-demand human-AI collaborative innovation tasks, including user interface and user experience design [46] and creative writing [64].

For more general collaborative tasks between humans and AI, systematic thinking provides a structured approach that benefits complex human-AI interactions. For instance, researchers developed CoQuest to facilitate interdisciplinary research and investigated the role of systematic thinking in interaction designs [49]. Similarly, researchers have highlighted the importance of systematic support in complex medical decision-making, where human-centered AI assists experts in intermediate stages such as hypothesis generation and data gathering, rather than focusing solely on final decisions [78]. In broader scenarios, systematic thinking can enhance decision-making processes, improve task efficiency, and help humans and AI navigate complex problems.

## 7.3 Affecting Human Reliance and Self-Confidence Appropriately with Collaborative AI

In the formative study, we observed that designers' over-reliance on AI or inappropriate overconfidence negatively impacts the effects of DSE. In CoExploreDS, guided by design methodologies, AI generates more targeted suggestions. Designers not only increased their reliance on AI assistance in CoExploreDS but also boosted their self-confidence, leading to improved design outcomes.

When collaborating with AI, human reliance and confidence exhibit complex dynamics, depending on how designers construct and calibrate their own capability models and those of the AI [58]. For example, designers' self-confidence can increase when their opinions are validated, either through positive reinforcement from AI or by successfully identifying and excluding erroneously generated content [53]. In practice, factors such as individual cognition and AI interpretability can influence this dynamic balance. From the designer's perspective, prevalent overconfidence in human decision-making [39] may manifest as an "ownership bias", where designers demonstrate an excessive preference for their own ideas, potentially overlooking better alternatives suggested by others [59]. This heightened confidence may in turn raise the risk of decision-making errors [58]. From the AI's perspective, satisfactory results during the early stages of collaboration can help designers form a positive first impression, leading to greater reliance on the AI [58]. Without stronger interpretability, designers may overestimate AI's capabilities, which may lead to missing potential opportunities.

Therefore, appropriately influencing human reliance on AI and self-confidence involves finding the optimal middle ground where designers can confidently trust their judgment and ensure their self-assessment corresponds to the actual quality of the design ("High Confidence-High Quality" scenario in Section 6.3). Researchers have proposed several recommendations for future AI-assisted decision-making systems, including making confidence calibration a default setting, emphasizing the cost of decision errors, and encouraging designers to approach problems from the opposite perspective [50]. However, for designs involving complex decision-making and reasoning processes, decision-making errors cannot be directly quantified through metrics; instead, they are often implicitly reflected in the final or intermediate design outcomes. In light of this, we proposed two additional design considerations for future human-AI collaboration design tools:

- User-centered visualization of AI "black-box" generation. This involves organizing AI-generated results in a way that is comprehensible to users, presenting them through intuitive interaction methods and logical content structures. The reasoning process should be broken down and visualized step-by-step, enabling users to build an accurate AI capability model and enhancing interpretability.
- Establishment of an AI self-evaluation system. For instance, providing self-assessments of novelty, practicality, design quality, and other metrics alongside the generated results. Multi-agent collaboration could be introduced as a potential solution to enhance the quality of AI outputs while offering designers referenceable self-reported evaluations from AI.

## 7.4 Limitations and Future Work

The inherent complexity of product design has consistently made assisting designers in addressing intricate tasks a challenge. Our work cut in from a perspective informed by established design theories, constructing the design space for human-AI interaction in terms of problems and solutions. This representation encompasses key considerations such as target users, client requirements, cost constraints, aesthetics, and functionality. Previous studies have proposed other assistance approaches, such as leveraging knowledge graphs [48], providing case-based recommendations [37], and applying invention-theory principles [45]. Additionally, some research has explored visualizing semantic relationships within the existing design space to provide content-related suggestions [6, 42]. However, our work and these studies are all limited in scope for focusing on one specific approach. Future work could integrate multiple methods. For instance, we could incorporate semantic information into the design space map and visualize the logical relationships between nodes. This enhancement would enable both designers and AI to more deeply consider specific dimensions of the design solutions.

Considering the dynamic and chaotic nature of ideation, another aspect such systems should consider is minimizing cognitive load for designers and uncovering deeper unconscious information. For instance, several participants suggested enhancing the system's automation, including support for automatic convergence and rating by LLMs. We plan to incorporate these features in future systems to effectively assist designers in understanding their current exploration. Additionally, our system is limited to capturing only textual intentions, inadequately representing the multimodal dimensions of the DSE process [25, 44]. The design methodologies employed also fail to capture crucial aspects that exist solely in designers' minds, such as non-logical reasoning, intuitive insights, and trait knowledge. Future research needs to explore additional methods to understand and simulate this implicit information to support DSE, providing a more comprehensive explanation.

# 8 Conclusion

In this paper, we explored human-AI collaborative DSE with the guidance of problem-solution co-evolution and design reasoning methods. In the formative study, we demonstrated the potential of LLM as a human-like design partner in DSE, identified key design styles of human-AI collaboration, and proposed three key strategies. We then developed CoExploreDS which enabled designers to frame and advance the DSE process with the problem-solution co-evolution model and design reasoning methods. Our findings showed that CoExploreDS facilitated human-AI collaborative DSE, enhancing outcomes and creativity due to two factors: CoExploreDS facilitated systematic exploration with less effort and appropriately affected designers' reliance on AI and self-confidence in collaboration. We further discussed how AI assistance could liberate designers from consciously adhering to design methodologies, extend systematic thinking to other collaborative innovation tasks, and influence designers' self-confidence and reliance on AI in collaboration.

# Acknowledgments

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

# References

[1] S. Altavilla and E. Blanco. 2020. Are AI tools going to be the new designers? A taxonomy for measuring the level of automation of design activities. *Proceedings of the Design Society: DESIGN Conference* 1 (2020), 81–90. doi:10.1017/dsd.2020.286

[2] Linden J Ball and Bo T Christensen. 2019. Advancing an understanding of design cognition and design metacognition: Progress and prospects. *Design Studies* 65 (2019), 35–59. doi:10.1016/j.destud.2019.10.003

[3] Linden J Ball, Thomas C Ormerod, and Nicola J Morley. 2004. Spontaneous analogising in engineering design: a comparative analysis of experts and novices. *Design Studies* 25, 5 (2004), 495–508. doi:10.1016/j.destud.2004.05.004

[4] Michael Mose Biskjaer, Peter Dalsgaard, and Kim Halskov. 2014. A constraint-based understanding of design spaces. In *Proceedings of the 2014 Conference on Designing Interactive Systems* (Vancouver, BC, Canada) *(DIS '14)*. Association for Computing Machinery, New York, NY, USA, 453–462. doi:10.1145/2598510.2598533

[5] John Brooke et al. 1996. SUS-A Quick and Dirty Usability Scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[6] Alice Cai, Steven R Rick, Jennifer L Heyman, Yanxia Zhang, Alexandre Filipowicz, Matthew Hong, Matt Klenk, and Thomas Malone. 2023. DesignAID: Using Generative AI and Semantic Diversity for Design Inspiration. In *Proceedings of The ACM Collective Intelligence Conference* (Delft, Netherlands) *(CI '23)*. Association for Computing Machinery, New York, NY, USA, 1–11. doi:10.1145/3582269.3615596

[7] Chunlei Chai, Fei Cen, Weiyu Ruan, Cheng Yang, and Hongting Li. 2015. Behavioral analysis of analogical reasoning in design: Differences among designers with different expertise levels. *Design Studies* 36 (2015), 3–30. doi:10.1016/j.destud.2014.07.001

[8] Senthil K Chandrasegaran, Karthik Ramani, Ram D Sriram, Imré Horváth, Alain Bernard, Ramy F Harik, and Wei Gao. 2013. The evolution, challenges, and future of knowledge representation in product design systems. *Computer-aided Design* 45, 2 (2013), 204–228. doi:10.1016/j.cad.2012.08.006

[9] Liuqing Chen, Zhaojun Jiang, Duowei Xia, Zebin Cai, Lingyun Sun, Peter Childs, and Haoyu Zuo. 2024. BIDTrainer: An LLMs-driven Education Tool for Enhancing the Understanding and Reasoning in Bio-inspired Design. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 676, 20 pages. doi:10.1145/3613904.3642887

[10] Liuqing Chen, Qianzhi Jing, Yixin Tsang, Qianyi Wang, Lingyun Sun, and Jianxi Luo. 2024. DesignFusion: Integrating Generative Models for Conceptual Design Enrichment. *Journal of Mechanical Design* 146, 11 (2024), 111703. doi:10.1115/1.4065487

[11] Liuqing Chen, Yaxuan Song, Shixian Ding, Lingyun Sun, Peter Childs, and Haoyu Zuo. 2024. TRIZ-GPT: An LLM-augmented method for problem-solving. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (Washington, DC, USA) *(International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. Volume 6: 36th International Conference on Design Theory and Methodology (DTM))*. American Society of Mechanical Engineers, New York, NY, USA, V006T06A010. doi:10.1115/DETC2024-143163

[12] Erin Cherry and Celine Latulipe. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Trans. Comput.-Hum. Interact.* 21, 4, Article 21 (June 2014), 25 pages. doi:10.1145/2617588

[13] Han Hee Choi and Mi Jeong Kim. 2017. The effects of analogical and metaphorical reasoning on design thinking. *Thinking skills and Creativity* 23 (2017), 29–41. doi:10.1016/j.tsc.2016.11.004

[14] Claus L Cramer-Petersen, Bo T Christensen, and Saeema Ahmed-Kristensen. 2019. Empirically analysing design reasoning patterns: Abductive-deductive reasoning patterns dominate design idea generation. *Design Studies* 60 (2019), 39–70. doi:10.1016/j.destud.2018.10.001

[15] Nathan Crilly. 2021. 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 (2021), 309–332. doi:10.1016/j.sheji.2021.07.003

[16] Nigel Cross. 2004. Expertise in design: an overview. *Design Studies* 25, 5 (2004), 427–441. doi:10.1016/j.destud.2004.06.002

[17] Richard Lee Davis, Thiemo Wambsganss, Wei Jiang, Kevin Gonyop Kim, Tanja Käser, and Pierre Dillenbourg. 2024. Fashioning Creative Expertise with Generative AI: Graphical Interfaces for Design Space Exploration Better Support Ideation Than Text Prompts. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 167, 26 pages. doi:10.1145/3613904.3642908

[18] H Onan Demirel, Molly H Goldstein, Xingang Li, and Zhenghui Sha. 2024. Human-centered generative design framework: an early design framework to support concept creation and evaluation. *International Journal of Human–Computer Interaction* 40, 4 (2024), 933–944. doi:10.1080/10447318.2023.2171489

[19] C Nathan DeWall, Roy F Baumeister, and EJ Masicampo. 2008. Evidence that logical reasoning depends on conscious processing. *Consciousness and Cognition* 17, 3 (2008), 628–645. doi:10.1016/j.concog.2007.12.004

[20] Andy Dong, Dan Lovallo, and Ronny Mounarath. 2015. The effect of abductive reasoning on concept selection decisions. *Design Studies* 37 (2015), 37–58. doi:10.1016/j.destud.2014.12.004

[21] Kees Dorst. 1995. Analysing design activity: new directions in protocol analysis. *Design Studies* 2, 16 (1995), 139–142.

[22] Kees Dorst. 2011. The core of 'design thinking'and its application. *Design Studies* 32, 6 (2011), 521–532. doi:10.1016/j.destud.2011.07.006

[23] Kees Dorst. 2019. Co-evolution and emergence in design. *Design Studies* 65 (2019), 60–77. doi:10.1016/j.destud.2019.10.005

[24] Kees Dorst and Nigel Cross. 2001. Creativity in the design process: co-evolution of problem–solution. *Design Studies* 22, 5 (2001), 425–437. doi:10.1016/S0142-694X(01)00009-6

[25] Tobias Drey, Jan Gugenheimer, Julian Karlbauer, Maximilian Milo, and Enrico Rukzio. 2020. VRSketchIn: Exploring the Design Space of Pen and Tablet Interaction for 3D Sketching in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3313831.3376628

[26] Martin J Eppler. 2006. A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing. *Information visualization* 5, 3 (2006), 202–210. doi:10.1057/palgrave.ivs.9500131

[27] Stefan Feuerriegel, Jochen Hartmann, Christian Janiesch, and Patrick Zschech. 2024. Generative AI. *Business & Information Systems Engineering* 66, 1 (2024), 111–126. doi:10.2139/ssrn.4443189

[28] Xiaoyun Fu, Haoyu Zhang, Liting Jing, Xiaoyan Fan, Congda Lu, and Shaofei Jiang. 2024. A constraint-driven conceptual design approach for product based on function-behavior-structure design process. *Computers & Industrial Engineering* 189 (2024), 109994. doi:10.1016/j.cie.2024.109994

[29] John S Gero, Udo Kannengiesser, and Nathan Crilly. 2022. Abstracting and formalising the design co-evolution model. *Design Science* 8 (2022), e14. doi:10.1017/dsj.2022.10

[30] Frederic Gmeiner, Humphrey Yang, Lining Yao, Kenneth Holstein, and Nikolas Martelaro. 2023. Exploring Challenges and Opportunities to Support Designers in Learning to Co-create with AI-based Manufacturing Design Tools. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 226, 20 pages. doi:10.1145/3544548.3580999

[31] Joachim Günther and Klaus Ehrlenspiel. 1999. Comparing designers from practice and designers with systematic design education. *Design Studies* 20, 5 (1999), 439–451. doi:10.1016/S0142-694X(99)00019-8

[32] Kim Halskov and Caroline Lundqvist. 2021. Filtering and Informing the Design Space: Towards Design-Space Thinking. *ACM Trans. Comput.-Hum. Interact.* 28, 1, Article 8 (Jan. 2021), 28 pages. doi:10.1145/3434462

[33] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, Amsterdam, Netherlands, 139–183.

[34] Yihan Hou, Manling Yang, Hao Cui, Lei Wang, Jie Xu, and Wei Zeng. 2024. C2Ideas: Supporting Creative Interior Color Design Ideation with a Large Language Model. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 172, 18 pages. doi:10.1145/3613904.3642224

[35] Xin Hu, Yu Tian, Keisuke Nagato, Masayuki Nakao, and Ang Liu. 2023. Opportunities and challenges of ChatGPT for design knowledge management. *Procedia CIRP* 119 (2023), 21–28. doi:10.1016/j.procir.2023.05.001

[36] Dongwook Hwang and Woojin Park. 2018. Design heuristics set for X: A design aid for assistive product concept generation. *Design Studies* 58 (2018), 89–126. doi:10.1016/j.destud.2018.04.003

[37] Lizhen Jia, Qingjin Peng, Runhua Tan, and Xuehong Zhu. 2019. Analogical stimuli retrieval approach based on R-SBF ontology model. *Journal of Engineering Design* 30, 10-12 (2019), 599–624. doi:10.1080/09544828.2019.1643830

[38] Shuo Jiang and Jianxi Luo. 2024. Autotriz: Artificial ideation with triz and large language models. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 88377. American Society of Mechanical Engineers, New York, NY, USA, V03BT03A055. doi:10.1115/DETC2024-143166

[39] Dominic DP Johnson and James H Fowler. 2011. The evolution of overconfidence. *Nature* 477, 7364 (2011), 317–320. doi:10.1038/nature10384

[40] David Jones, James Gopsill, Ric Real, Chris Snider, Harry Felton, Lee Kent, Mark Goudswaard, Owen Freeman Gebler, and Ben Hicks. 2024. The prototype taxonomised: Towards the capture, curation, and integration of physical models in new product development. *Computers in Industry* 155 (2024), 104059. doi:10.1016/j.compind.2023.104059

[41] Majeed Kazemitabaar, Jack Williams, Ian Drosos, Tovi Grossman, Austin Zachary Henley, Carina Negreanu, and Advait Sarkar. 2024. Improving Steering and Verification in AI-Assisted Data Analysis with Interactive Task Decomposition. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 92, 19 pages. doi:10.1145/3654777.3676345

[42] Jingoog Kim and Mary Lou Maher. 2023. The effect of AI-based inspiration on human design ideation. *International Journal of Design Creativity and Innovation* 11, 2 (2023), 81–98. doi:10.1080/21650349.2023.2167124

[43] Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners, In Advances in Neural Information Processing Systems. *Advances in neural information processing systems* 35, 22199–22213. https://proceedings.neurips.cc/paper_files/paper/2022/file/8bb0d291acd4acf06ef112099c16f326-Paper-Conference.pdf

[44] Tobias Langlotz, Jonathan Sutton, and Holger Regenbrecht. 2024. A Design Space for Vision Augmentations and Augmented Human Perception using Digital Eyewear. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 966, 16 pages. doi:10.1145/3613904.3642380

[45] CKM Lee, Jingying Liang, KL Yung, and KL Keung. 2024. Generating TRIZ-inspired guidelines for eco-design using Generative Artificial Intelligence. *Advanced Engineering Informatics* 62 (2024), 102846. doi:10.1016/j.aei.2024.102846

[46] Zhipeng Li, Christoph Gebhardt, Yves Inglin, Nicolas Steck, Paul Streli, and Christian Holz. 2024. SituationAdapt: Contextual UI Optimization in Mixed Reality with Situation Awareness via LLM Reasoning. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 43, 13 pages. doi:10.1145/3654777.3676470

[47] Ang Liu, Yuchen Wang, Ivan Teo, and Stephen Lu. 2019. Constraint management for concept ideation in conceptual design. *CIRP Journal of Manufacturing Science and Technology* 24 (2019), 35–48. doi:10.1016/j.cirpj.2018.11.006

[48] Qiyu Liu, Kai Wang, Yan Li, and Ying Liu. 2020. Data-driven concept network for inspiring designers' idea generation. *Journal of Computing and Information Science in Engineering* 20, 3 (2020), 031004. doi:10.1115/1.4046207

[49] Yiren Liu, Si Chen, Haocong Cheng, Mengxia Yu, Xiao Ran, Andrew Mo, Yiliu Tang, and Yun Huang. 2024. How AI Processing Delays Foster Creativity: Exploring Research Question Co-Creation with an LLM-based Agent. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 17, 25 pages. doi:10.1145/3613904.3642698

[50] Shuai Ma, Xinru Wang, Ying Lei, Chuhan Shi, Ming Yin, and Xiaojuan Ma. 2024. "Are You Really Sure?" Understanding the Effects of Human Self-Confidence Calibration in AI-Assisted Decision Making. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 840, 20 pages. doi:10.1145/3613904.3642671

[51] Benjamin Matthews, Barrie Shannon, and Mark Roxburgh. 2023. Destroy All Humans: The Dematerialisation of the Designer in an Age of Automation and its Impact on Graphic Design—A Literature Review. *International Journal of Art & Design Education* 42, 3 (2023), 367–383. doi:10.1111/jade.12460

[52] Yakira Mirabito and Kosa Goucher-Lambert. 2021. Connecting design actions, reasoning, and outcomes in concept-selection. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 85420. American Society of Mechanical Engineers, New York, NY, USA, V006T06A037. doi:10.1115/DETC2021-71830

[53] Don A Moore and Paul J Healy. 2008. The Trouble with Overconfidence. *Psychological review* 115, 2 (2008), 502.

[54] Gjoko Muratovski, Ken Friedman, Don Norman, and Steven Heller. 2022. *Research for Designers: A Guide to Methods and Practice (2nd Edition) - Sample Text*. SAGE Publications Ltd, Thousand Oaks, USA.

[55] Adnan Niazi, Jian S. Dai, Stavroula Balabani, and Lakmal Seneviratne. 2005. Product Cost Estimation: Technique Classification and Methodology Review. *Journal of Manufacturing Science and Engineering* 128, 2 (2005), 563–575. doi:10.1115/1.2137750

[56] Jordan Nickel, P Robert Duimering, and Ada Hurst. 2022. Manipulating the design space to resolve trade-offs: Theory and evidence. *Design Studies* 79 (2022), 101095. doi:10.1016/j.destud.2022.101095

[57] Bernard A. Nijstad and Wolfgang Stroebe. 2006. How the Group Affects the Mind: A Cognitive Model of Idea Generation in Groups. *Personality and Social Psychology Review* 10, 3 (2006), 186–213. doi:10.1207/s15327957pspr1003_1

[58] Mahsan Nourani, Chiradeep Roy, Jeremy E Block, Donald R Honeycutt, Tahrima Rahman, Eric Ragan, and Vibhav Gogate. 2021. Anchoring Bias Affects Mental Model Formation and User Reliance in Explainable AI Systems. In *Proceedings of the 26th International Conference on Intelligent User Interfaces* (College Station, TX, USA) *(IUI '21)*. Association for Computing Machinery, New York, NY, USA, 340–350. doi:10.1145/3397481.3450639

[59] Balder Onarheim and Bo T Christensen. 2012. Distributed idea screening in stage–gate development processes. *Journal of Engineering Design* 23, 9 (2012), 660–673. doi:10.1080/09544828.2011.649426

[60] Jonas Oppenlaender, Rhema Linder, and Johanna Silvennoinen. 2024. Prompting AI art: An Investigation into the Creative Skill of Prompt Engineering. *International Journal of Human–Computer Interaction* 0, 0 (2024), 1–23. doi:10.1080/10447318.2024.2431761

[61] Ozgu Ozkan and Fehmi Dogan. 2013. Cognitive strategies of analogical reasoning in design: Differences between expert and novice designers. *Design Studies* 34, 2 (2013), 161–192. doi:10.1016/j.destud.2012.11.006

[62] Prabir Sarkar and Amaresh Chakrabarti. 2011. Assessing design creativity. *Design Studies* 32, 4 (2011), 348–383. doi:10.1016/j.destud.2011.01.002

[63] Kihoon Son, DaEun Choi, Tae Soo Kim, Young-Ho Kim, and Juho Kim. 2024. Genquery: Supporting expressive visual search with generative models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 180, 19 pages. doi:10.1145/3613904.3642847

[64] Sangho Suh, Meng Chen, Bryan Min, Toby Jia-Jun Li, and Haijun Xia. 2024. Luminate: Structured Generation and Exploration of Design Space with Large Language Models for Human-AI Co-Creation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 644, 26 pages. doi:10.1145/3613904.3642400

[65] Toshiharu Taura and Yukari Nagai. 2017. Creativity in Innovation Design: the roles of intuition, synthesis, and hypothesis. *International Journal of Design Creativity and Innovation* 5, 3-4 (2017), 131–148. doi:10.1080/21650349.2017.1313132

[66] Karl T Ulrich and Steven D Eppinger. 2016. *Product design and development*. McGraw-hill, New York, NY, USA.

[67] Maarten Van Someren, Yvonne F Barnard, and J Sandberg. 1994. The think aloud method: a practical approach to modelling cognitive. *London: AcademicPress* 11, 6 (1994), 1–205.

[68] Willemien Visser. 1994. Organisation of design activities: opportunistic, with hierarchical episodes. *Interacting with computers* 6, 3 (1994), 239–274.

[69] Boheng Wang, Haoyu Zuo, Zebin Cai, Yuan Yin, Peter Childs, Lingyun Sun, and Liuqing Chen. 2023. A Task-Decomposed AI-Aided Approach for Generative Conceptual Design. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 87349. American Society of Mechanical Engineers, New York, NY, USA, V006T06A009. doi:10.1115/DETC2023-109087

[70] Ru Wang, Yanshao Sun, Tao Peng, Yiwei Hua, Guoxin Wang, and Yan Yan. 2024. A knowledge graph-aided decision guidance method for product conceptual design. *Journal of Engineering Design* (2024), 1–40. doi:10.1080/09544828.2024.2368405

[71] Sitong Wang, Savvas Petridis, Taeahn Kwon, Xiaojuan Ma, and Lydia B Chilton. 2023. PopBlends: Strategies for Conceptual Blending with Large Language Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 435, 19 pages. doi:10.1145/3544548.3580948

[72] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting

elicits reasoning in large language models, In Advances in Neural Information Processing Systems. *Advances in neural information processing systems* 35, 24824–24837. https://proceedings.neurips.cc/paper_files/paper/2022/file/9d5609613524ecf4f15af0f7b31abca4-Paper-Conference.pdf

[73] Stefan Wiltschnig, Bo T Christensen, and Linden J Ball. 2013. Collaborative problem–solution co-evolution in creative design. *Design Studies* 34, 5 (2013), 515–542. doi:10.1016/j.destud.2013.01.002

[74] Robert F Woodbury and Andrew L Burrow. 2006. Whither design space? *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 20, 2 (2006), 63–82. doi:10.1017/S0890060406060057

[75] Zhuohao Wu, Danwen Ji, Kaiwen Yu, Xianxu Zeng, Dingming Wu, and Mohammad Shidujaman. 2021. AI creativity and the human-AI co-creation model. In *Human-Computer Interaction. Theory, Methods and Tools: Thematic Area, HCI 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part I 23.* Springer, New York, NY, USA, 171–190. doi:10.1007/978-3-030-78462-1_13

[76] J.D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 437, 21 pages. doi:10.1145/3544548.3581388

[77] Guanglu Zhang, Ayush Raina, Jonathan Cagan, and Christopher McComb. 2021. A cautionary tale about the impact of AI on human design teams. *Design Studies* 72 (2021), 100990. doi:10.1016/j.destud.2021.100990

[78] Shao Zhang, Jianing Yu, Xuhai Xu, Changchang Yin, Yuxuan Lu, Bingsheng Yao, Melanie Tory, Lace M. Padilla, Jeffrey Caterino, Ping Zhang, and Dakuo Wang. 2024. Rethinking Human-AI Collaboration in Complex Medical Decision Making: A Case Study in Sepsis Diagnosis. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 445, 18 pages. doi:10.1145/3613904.3642343

[79] Chengbo Zheng, Yuanhao Zhang, Zeyu Huang, Chuhan Shi, Minrui Xu, and Xiaojuan Ma. 2024. DiscipLink: Unfolding Interdisciplinary Information Seeking Process via Human-AI Co-Exploration. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 91, 20 pages. doi:10.1145/3654777.3676366

[80] Daixing Zhong, Jiashuang Fan, Gangjun Yang, Baozhen Tian, and Yukun Zhang. 2022. Knowledge management of product design: A requirements-oriented knowledge management framework based on Kansei engineering and knowledge map. *Advanced Engineering Informatics* 52 (2022), 101541. doi:10.1016/j.aei.2022.101541

[81] Ruican Zhong, Donghoon Shin, Rosemary Meza, Predrag Klasnja, Lucas Colusso, and Gary Hsieh. 2024. AI-Assisted Causal Pathway Diagram for Human-Centered Design. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 2, 19 pages. doi:10.1145/3613904.3642179

[82] Jiayi Zhou, Renzhong Li, Junxiu Tang, Tan Tang, Haotian Li, Weiwei Cui, and Yingcai Wu. 2024. Understanding Nonlinear Collaboration between Human and AI Agents: A Co-design Framework for Creative Design. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 170, 16 pages. doi:10.1145/3613904.3642812

# A Details of Formative Study

## A.1 Calculation of Design Quality

$U = L \times R \times F$, $U$ is calculated as the product of the level of importance ($L$), the popularity of product usage ($R$), and its frequency ($F$). $L$ is assessed using a 5-point scale, while $R$ and $F$ are measured on a scale from 0 to 1, precise to one decimal place. We performed Kendall's W consistency test and linear normalization on the scores for $N$ and $U$. Subsequently, $U$ was converted to a 1-to-7 scale to facilitate comparison with $N$ and to have equal weight in calculating the sum score of design quality.

## A.2 Rating Forms for Experts

The expert rating forms included the following components:

**Final Design Problem**: The dimensions of product design considered by the participants during the design process, such as "How to adapt to different environments?" and "How to design the external structure to meet specific functions?"

**Final Design Solution**: A textual description of the final design concept produced by the participants.

**Novelty**: The innovative and unusual features of the product compared to existing products. For products with multiple functions, each function should be compared individually with the corresponding functions of existing products.

**Usefulness**: A product's usefulness is measured in terms of its actual use:

- **Level of Importance**: Products that fulfill more fundamental human needs are assigned higher values for usefulness as defined by Maslow's hierarchy of needs.
- **Rate of Popularity of Usage**: This refers to the proportion of users who utilize the product within a specified period, measuring its popularity.
- **Frequency of Usage**: The measure of how often the product is used within a given period.

## A.3 Examples of Creation Outputs of Participants

Table 3 provides examples of the creation outputs and the associated scores in the formative study. The design task was to design a commuting electric scooter. To enhance clarity in visualizing how outcomes correlate with various scores, examples of sub-function scoring were provided. The evaluation criteria for overall solutions were identical to those for sub-functions.

**Table 3: Examples of the creation outputs of four participants and the associated average scores.**

| Metrics | Score | Examples From Four Participants |
|---|---|---|
| $N$ | 7.00 | The cover on the foot pedal can be opened to reveal a foldable seat, transforming it into an electric bicycle... |
| | 6.67 | Automatically adjusts the scooter's center of gravity based on road conditions and speed, using sensors... |
| | 5.67 | Features removable storage modules of varying sizes integrated into the frame... |
| | 5.00 | Equipped with a find-my-scooter feature, using built-in bells or flashing lights to help users locate it... |
| | 4.33 | Fitted with shock-absorbing tires... |
| | 3.67 | Equipped with a smartphone holder... |
| $U$ | 7.00 | The front fog light can be detached and used as a flashlight... |
| | 4.10 | The scooter's weight is controlled to be under 10 kilograms, and it can be folded to fit into a backpack... |
| | 3.10 | The cover on the foot pedal can be opened to reveal a foldable seat, transforming it into an electric bicycle... |
| | 2.36 | Equipped with a find-my-scooter feature... |
| | 1.21 | It features unique hub patterns and interchangeable colors... |
| | 1.00 | The handlebars, deck, and stem are all foldable... |

## A.4 Design Reasoning Methods Observed in DSE

Table 4 illustrates the different design reasoning methods used by both designers and AI during the human-AI collaborative DSE process. It shows how often each reasoning method—deductive, inductive, abductive, and analogical—was applied, along with the specific counts for problems and solutions in both roles. Unrecorded co-evolution episodes primarily involved simple queries or content lacking clear reasoning methods.

**Table 4: Design reasoning methods in human-AI collaborative exploration**

| Role | Reasoning | Total | Problems | Solutions |
|---|---|---|---|---|
| Designers | Deductive | 17 | 7 | 10 |
| | Inductive | 24 | 18 | 6 |
| | Abductive | 58 | 44 | 14 |
| | Analogical | 14 | 5 | 9 |
| AI | Deductive | 54 | 13 | 41 |
| | Inductive | 52 | 33 | 19 |
| | Abductive | 75 | 40 | 35 |
| | Analogical | 0 | 0 | 0 |

# B Details of CoExploreDS

## B.1 Algorithm of Design Map Visualization

The method for constructing the design space visualization map is shown in Algorithm 1. The inputs of the method include the root node $r$ (descriptions of the design task), and the set of all edges $\mathcal{E}$. Each edge is defined by a list that includes two connected nodes (with details such as node ID, type, and inference method) and the connection type (either user-defined or automatic). The root node is usually the source problem node. The algorithm outputs both an adjacency graph $\mathcal{G}$ and a level-layered graph $\mathcal{N}_l$. The adjacency graph lists adjacent nodes for each node, while the level-layered graph organizes nodes by levels. For visualization, node positions are determined from the level-layered graph, and connections are drawn based on the adjacency graph. Finally, get an image of the design space map $IMG_{\mathcal{G}}$. The design space map construction algorithm first establishes adjacency for all nodes, then builds the level-layered graph using an improved Breadth-First Search algorithm.

**Algorithm 1:** Design Space Visualization Map Construction

**Input:** The root node $r$, the set of all edges $\mathcal{E}$
**Output:** Adjacency graph $\mathcal{G}$, level-layered graph of nodes $\mathcal{N}_l$, visualization image of design space map $IMG_{\mathcal{G}}$

$\mathcal{G} \leftarrow Dict(List)$, $\mathcal{N}_l \leftarrow Dict(List)$
**for** $e$ *in* $\mathcal{E}$ **do**
  $\mathcal{G}.append(e)$ // Add nodes information of $\mathcal{E}$ into $\mathcal{G}$
$q \leftarrow Deque([r,0])$ // $q$ is the deque of node and its level
$\mathcal{N}_v \leftarrow Set()$ // $\mathcal{N}_v$ is the set of nodes have been visited
**while** $q$ **do**
  $n, l = q.popfront()$ // $n$ is the node of an element of $q$, $l$ is the level of $n$
  **if** $n$ *not in* $\mathcal{N}_v$ **then**
    *Add* $n$ *to* $\mathcal{N}_v$ *and* $\mathcal{N}_l[l]$ // Mark that $n$ has been traversed and at $l$ level
  **if** *neighbours of* $n$ *not in* $\mathcal{N}_v$ **then**
    *Add neighbours of* $n$ *into* $q$
$IMG_{\mathcal{G}} \leftarrow PaintImage(\mathcal{G}, \mathcal{N}_l)$
**return** $\mathcal{G}$, $\mathcal{N}_l$, $IMG_{\mathcal{G}}$

## B.2 Algorithm of Proactive Suggestions Generation

According to Algorithm 2, the input includes the position of the selected node $n$, as well as the reasoning methods they used $\mathcal{M}$, along with the current adjacency graph $\mathcal{G}$ and the level-layered graph $\mathcal{N}_l$ which contains the type of problem or solution for both the node and its parent and sibling nodes. The algorithm outputs a suggestion information $\mathcal{L}_{\mathcal{I}}$.

**Algorithm 2:** Proactive Suggestions Generation Based on Design Space Maps.

**Input:** The node selected $n$, the method of adding a node $\mathcal{M}$, the design space map $\mathcal{G}$, level-layered nodes $\mathcal{N}_l$
**Output:** A dict list of suggestion information $\mathcal{L}_{\mathcal{I}}$

$\mathcal{L}_{\mathcal{I}}$ = List(dict)
$\_l \leftarrow is\_Linear(\mathcal{G}, \mathcal{N}_l, n)$
$\_d \leftarrow is\_Divergent(\mathcal{G}, \mathcal{N}_l, n)$
$\_a \leftarrow is\_Associative(\mathcal{G}, \mathcal{N}_l, n)$
// Judge whether the node is linear, divergent, or associative.
$\mathcal{L}_p \leftarrow calculatePosition(\mathcal{G}, n, \mathcal{N}_l, \_l, \_d, \_a)$ // Get position information list of generated nodes $\mathcal{L}_p$
$\mathcal{D}_p \leftarrow ListToDictByPosition(\mathcal{L}_p)$ // Construction directory of position and nodes number at $p$
**for** $\mathcal{I}_p$ *in* $\mathcal{D}_p$ **do**
  $N \leftarrow \mathcal{D}_p[\mathcal{I}_p]$ // Get number of generated nodes at $\mathcal{I}_p$
  $\mathcal{I}_t \leftarrow calculateType(\mathcal{I}_p, N)$ // Get types of generated nodes at $p$
  **if** $\mathcal{M}$ *is 'Auto'* **then**
    $\mathcal{I}_m \leftarrow calculateMethod(\mathcal{G}, \mathcal{N}_l, \mathcal{I}_t)$ // Get methods of generated nodes at $p$
  **else**
    $\mathcal{I}_m \leftarrow$ [repeat $\mathcal{M}$ for $N$ times]
  $\mathcal{L}_{\mathcal{I}}$.append({'Pos': $p$, 'Type': $\mathcal{I}_t$, 'Method': $\mathcal{I}_m$})
**return** $\mathcal{L}_{\mathcal{I}}$

# C Details of User Study

## C.1 Demographics

Table 5 the demographic information of participants in the user study. Group A used the CoExploreDS during the design tasks, while Group B used the baseline system. The sample size for the user study was determined based on similar studies in the field of HCI [41, 64, 79].

## C.2 Baseline System

Figure 11 shows the interface of the baseline system used for the user study. The baseline system used was a simplified version of CoExploreDS, retaining only the mind map function and basic AI generation capabilities, with the "co-evolution" and "reasoning patterns" functionalities removed. In the baseline system, the text content of a user-selected node was used as a prompt to retrieve GPT-generated responses via an API, which were then displayed on the canvas in a new node.

## C.3 Interview Questions

Participants in Group A were asked the following questions after completing tasks using CoExploreDS:

**Table 5: Demographics information of participants.**

| Group | PID | Age | Gender | Experience |
|---|---|---|---|---|
| A | P1 | 24 | Female | 3-5 years |
| | P2 | 24 | Female | 3-5 years |
| | P3 | 25 | Female | >5 years |
| | P4 | 25 | Female | 3-5 years |
| | P5 | 24 | Male | 3-5 years |
| | P6 | 24 | Female | 3-5 years |
| | P7 | 25 | Male | 3-5 years |
| | P8 | 25 | Male | >5 years |
| | P9 | 24 | Female | 3-5 years |
| | P10 | 24 | Female | 3-5 years |
| | P11 | 24 | Female | 3-5 years |
| | P12 | 24 | Male | 3-5 years |
| | P13 | 22 | Female | 3 years |
| | P14 | 22 | Female | 3 years |
| | P15 | 23 | Female | 3-5 years |
| | P16 | 24 | Female | 3-5 years |
| B | P17 | 24 | Male | 3-5 years |
| | P18 | 24 | Female | 3-5 years |
| | P19 | 24 | Female | 3-5 years |
| | P20 | 24 | Male | 3-5 years |
| | P21 | 29 | Female | >5 years |
| | P22 | 23 | Female | 3 years |
| | P23 | 23 | Female | 3-5 years |
| | P24 | 23 | Female | 3-5 years |
| | P25 | 23 | Female | 3-5 years |
| | P26 | 23 | Male | 3-5 years |
| | P27 | 25 | Male | >5 years |
| | P28 | 23 | Female | 3-5 years |
| | P29 | 24 | Female | 3-5 years |
| | P30 | 24 | Female | 3 years |
| | P31 | 23 | Male | 3-5 years |
| | P32 | 23 | Female | 3-5 years |

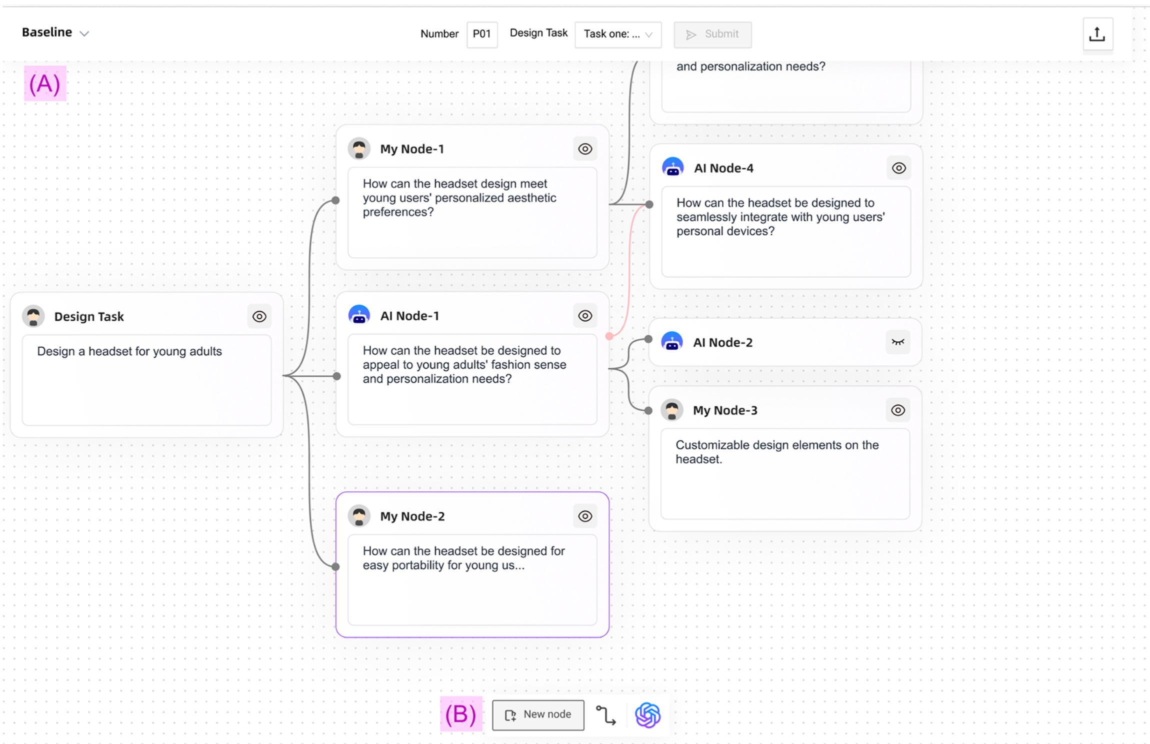


**Figure 11: Interface of the baseline system. (A) The main canvas is the same as in CoExploreDS. (B) The bottom toolbar was a simplified version of CoExploreDS with the "co-evolution" and "reasoning patterns" functionalities removed.**

(1) What assistance did the suggestions in the Quick Assist panel provide? Did it meet your expectations? How did it impact your design process and the final solution?
(2) What assistance did the system provide when using the AI generation feature in the bottom toolbar? Was it as expected? How did it affect your design process and the final solution?
(3) Does CoExploreDS meet your expectations? How would you like to see it improved?
(4) Discuss the strengths and weaknesses of this system.
(5) Reflecting on your design process, what is your general approach to design thinking?
(6) Which reasoning methods do you excel at, and which do you struggle with?
(7) What do you think AI is better at? In which of its reasoning methods are you more confident?
(8) What role does AI play in your ideation and reasoning process?

Participants in Group B were asked the following questions after completing tasks using the baseline:

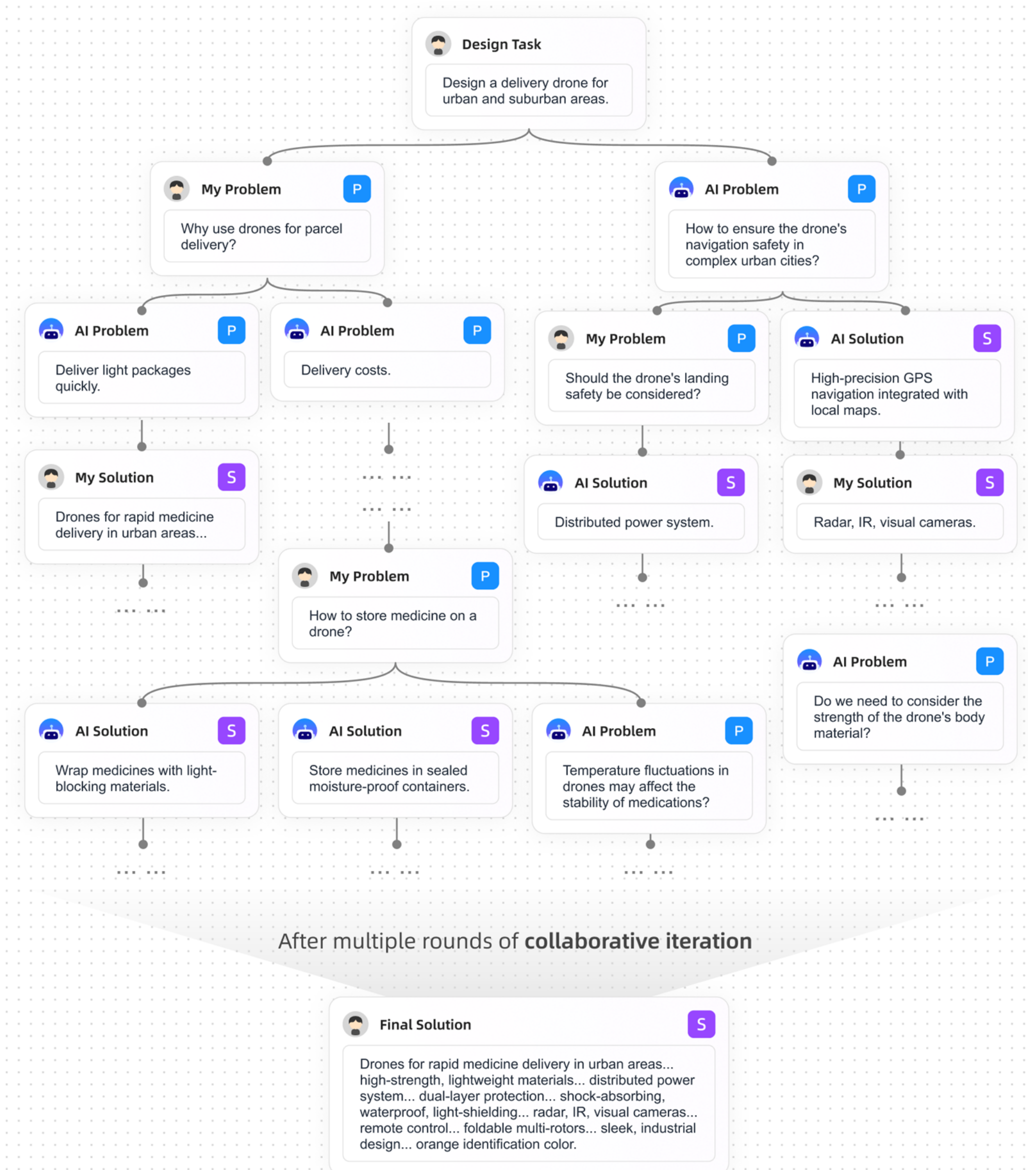


**Figure 12: Example of P16's design process with CoExploreDS.**

(1) What assistance did the system provide when using the AI generation feature in the bottom toolbar? Was it as expected? How did it affect your design process and the final solution?
(2) Discuss the strengths and weaknesses of this system.
(3) Reflecting on your design process, what is your general approach to design thinking?
(4) Which reasoning methods do you excel at, and which do you struggle with?
(5) What do you think AI is better at? In which of its reasoning methods are you more confident?
(6) What role does AI play in your ideation and reasoning process?

## C.4 Self-Customized Questionnaires

A seven-point custom questionnaire for collaborative attitudes (RQ2):

(1) In the design process, I rely on AI.
(2) In the design process, I am confident in my results.

## C.5 Example of Design Process with CoExploreDS

Figure 12 illustrates an example of the DSE process with CoExploreDS.

Received 12 September 2024; revised 10 December 2024; accepted 16 January 2025