Design Science

How generative AI supports human in conceptual design

Liuqing Chen [1], Yaxuan Song [1], Jia Guo[1], Lingyun Sun [1], Peter Childs[2] and Yuan Yin[2]

[1]*College of Computer Science and Technology, Zhejiang University, Zhejiang, China*
[2]*Dyson School of Design Engineering, Imperial College London, London, UK*

Abstract

Generative Artificial Intelligence (Generative AI) is a collection of AI technologies that can generate new information such as texts and images. With its strong capabilities, Generative AI has been actively studied in creative design processes. However, limited studies have explored the roles of humans and Generative AI in conceptual design processes, which leaves a gap for human–AI collaboration investigation. To address this gap, this study attempts to uncover the contributions of different Generative AI technologies in assisting humans in the conceptual design process. Novice designers were recruited to complete two design tasks in the condition of with or without the assistance of Generative AI. The results revealed that Generative AI primarily assists humans in the problem definition and idea generation stages, while the idea selection and evaluation stage remains predominantly human-led. Additionally, with the assistance of Generative AI, the idea selection and evaluation stages were further enhanced. Based on the findings, we discussed the role of Generative AI in human–AI collaboration and the implications for enhancing future conceptual design support with Generative AI's assistance.

1. Introduction

Conceptual design, which translates design requirements into preliminary design solutions, is a crucial phase in the product design process (French et al., 1985). However, exploring problems and generating design solutions place high demands on designers' knowledge and reasoning abilities (Myrup Andreasen et al., 2015), reflecting the complex nature of conceptual design. Various design theories and methodologies have been proposed to help designers gain a more comprehensive understanding of the conceptual design process and to assist designers in developing creative ideas and solutions, such as TRIZ theory (Al'tshuller, 1999), FBS model (Gero and Kannengiesser, 2014) and C-K theory (Hatchuel and Weil, 2009). These theories have made the representation of the conceptual design process more structured. However, the effective application of these methodologies still depends on designers' own knowledge and experience, which pose significant challenges for novice designers. With ongoing technological progress, some computational methods and tools have been proposed to alleviate novice designers' cognitive burden (Sarica and Luo, 2024; Cantamessa et al., 2020). For example, semantic networks (Luo et al., 2019; Shi et al., 2017) and case databases (Robles et al., 2009; Deldin and Schuknecht, 2013) have been

Received 13 August 2024
Revised 25 January 2025
Accepted 29 January 2025

Corresponding author
Yuan Yin
y.yin19@imperial.ac.uk

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Des. Sci., vol. 11, e9
journals.cambridge.org/dsj
DOI: 10.1017/dsj.2025.2

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press

established to support designers during the conceptual design stages. Although these tools provide inspiring stimuli, they do not offer corresponding solution suggestions for the specific design situation encountered. This means that designers still need to reason from the case domain to the problem domain to generate concrete solutions.

Driven by technological advancements in machine learning, various Generative AI models including transformer (Vaswani et al., 2017), diffusion models (Ho et al., 2020) and GANs (Generative Adversarial Networks) (Karras et al., 2019) have demonstrated significant potential and powerful performances. Building on these technologies, applications such as ChatGPT, Stable diffusion and Midjourney are making Generative AI more accessible and easy to use for consumers. Within the realm of design, text-to-text models and text-to-image models have attracted unique attention due to their ability to integrate seamlessly with the creative process and enhance the efficiency of design iterations. These models have become the most widely utilized generative techniques in research that combine Generative AI with design processes (Wu et al., 2024). Specifically, researchers have explored the application of text-to-text models in guiding the design process (Chen et al., 2024b), assisting with divergent and convergent thinking (Wang et al., 2023) and generating innovative solutions (Zhu and Luo, 2023), holding great potential for creativity enhancement in the innovation process (Sarica and Luo, 2024). For text-to-image technologies, they help designers visualize their design ideas quickly and reduce the time and skill demand of manual sketching for human designers (Choi et al., 2024). Also, text-to-image technologies can generate visual stimuli for design ideation based on user-input text prompts (Liu et al., 2023; Wadinambiarachchi et al., 2024).

Although researchers have recognized the importance of Generative AI in the conceptual design process, there is still a lack of empirical evidence for the effect of Generative AI in different stages of conceptual design. This gap may impede researchers from reflecting on and improving the developed collaborative tool designs. To fill the research gap, this study attempts to explore how Generative AI assists humans in conceptual design processes. Specifically, we recruited four groups of participants to finish two design tasks with (or without) the assistance of Generative AI (ChatGPT or Midjourney). We assessed human–AI collaboration in the conceptual design process through multiple dimensions, including the stages in which Generative AI helped designers, the stages led by humans, participants' assessments of the Generative AI tool's performance, expert ratings of the design outputs and prompt analysis of the strategies utilized by human designers during the four stages of conceptual design. We found that Generative AI primarily assists humans in the problem definition and idea generation stages, while the idea selection and evaluation stage remains predominantly human-led. Additionally, with the assistance of Generative AI, the idea selection and evaluation stage was further enhanced.

Our study provides an empirical contribution to the Generative AI-powered creativity support research by illustrating how Generative AI supports humans in conceptual design at a stage level. It further elaborates on Generative AI's role in different stages across conceptual design. Further, we demonstrate implications for future conceptual design support under Generative AI's help.

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press

2. Literature review

2.1. Conceptual design

According to previous research, the product design process can be divided into four phases: analysis of a problem, conceptual design, embodiment of scheme and detailing (French et al., 1985). Among these, conceptual design, which encompasses preliminary decision-making and design concept generation, is regarded as the key part of the design process (Eppinger and Ulrich, 1995). A few conceptual design models have been proposed to explain the stages of conceptual design. For example, Goodman-Deane et al. (2016) outlined that the conceptual design process encompasses four stages: manage (deciding what actions to take next), explore (identifying needs), create (generating ideas) and evaluate (judging and testing the design concepts). Jasmine (2020) delineated the design process into several distinct phases: establishing design requirements, assessing technology availability, sketching concepts and layouts, performing analysis and making trade-offs, optimizing revisions and developing a preliminary design. Some researchers also promoted applying the conventional design process model to conceptual design, such as the double diamond design process (Design Council, 2019), which includes discover, define, develop and deliver. Building on previous frameworks and considering the integral role of Generative AI, this study defines the conceptual design process as consisting of four stages: problem definition, idea generation, idea selection and evaluation and idea evolution. This serves as the foundation for our experiment and underpins the research findings and conclusions presented in this study.

Although conceptual design is essential for design processes, it is challenging to obtain creative design ideas of high originality and novelty based on designers' own efforts. Many computer-aided conceptual design support methods and tools have been proposed to offer creativity support to designers. For example, knowledge- or heuristics-based stimulation approaches can retrieve and mapping of source knowledge into the target design domain (Jiang et al., 2022). Some studies have attempted to utilize the information in patents, research papers, or encyclopedia data to construct semantic networks (Luo et al., 2019; Sarica et al., 2020). By computing the semantic distances between design goals and knowledge in a database, these methods could offer design stimuli or knowledge to human designers. However, these stimuli-based methods still require designers' cross-domain reasoning to complete the final design concept adapting to the current problem scenario.

2.2. Generative AI in conceptual design

Driven by technological advancements in machine learning, such as Generative Adversarial Networks (GANs) (Goodfellow et al., 2014), Variational Autoencoders (VAEs) (Kingma and Welling, 2014) and transformers (Vaswani et al., 2017), various Generative AI models including GPTs (Radford et al., 2018), BERT (Kenton and Toutanova, 2019) and StyleGAN (Karras et al., 2019) have demonstrated significant potential and powerful performance. Among these, text-to-text and text-to-image models have garnered considerable attention in the field of conceptual design (Wu et al., 2024), and have sparked a series of studies on how to smoothly integrate these two types of models into existing workflows (Mahdavi Goloujeh et al., 2024; Guo et al., 2024).

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press

Text-to-text tools, enhanced by Large Language Models (LLMs), such as ChatGPT, Llama and BERT, can generate natural and fluent answers to comprehend user input and provide contextual solutions in natural language. In the conceptual design domain, Generative AI-based text generation has been applied to requirement extraction (Shahin et al., 2024), creative ideation (Suh et al., 2024), solution generation (Chen et al., 2024c) and so on. For our experiments, as the experiment was carried out from May 2023 to June 2023, we specifically chose GPT-3.5 due to its robust capabilities in generating coherent and contextually relevant text outputs, which has been widely applied in Generative AI-assisted design research (Chen et al., 2024b; Chen et al., 2024a; Chen et al., 2024d).

Some text-to-image models have also been used to aid designers in early concept development by providing visual references and multimodal stimuli (Kwon et al., 2023). These models, such as Midjourney, DALL-E and Stable Diffusion, promote rapid exploration and iteration through visualization, enabling designers to better express their design concepts. Among these, Midjourney stands out both commercially and in terms of model performance, which has been widely adopted in human–AI collaboration research (Tan and Luhrs, 2024; Wadinambiarachchi et al., 2024; Mahdavi Goloujeh et al., 2024) due to its impressive image generation quality and user-friendly features.

Although various work has been done to develop Generative AI-based design tools and methodologies, there is still a lack of empirical evidence for the effects of Generative AI in conceptual design processes. Thus, in this study, we adapt experimental methods from traditional design research to explore the influence of two representative Generative AI models (i.e., text-to-text and text-to-image models) on different conceptual design stages, contributing new empirical evidence to the design community.

During the human–AI collaboration, Generative AI can assist human designers by generating concepts for selection, evaluation and iteration. Additionally, the output from Generative AI can inspire designers to develop more innovative ideas as the information can often expand designers' knowledge and exploration scope. These capabilities create unprecedented opportunities, particularly for novice designers, by significantly lowering the barriers to cross-disciplinary design and rapid visualization. Therefore, this study focuses on novice designers as research subjects to ensure more targeted research conclusions.

3. Experimental study

To gain deeper insights into the human–AI collaboration paradigm, a human–AI co-design study was conducted. Midjourney (a text-to-image Generative AI model) and GPT-3.5 (a text-to-text Generative AI model) were selected as example Generative AIs. The selection of these general-purpose Generative AI tools, rather than design-specific alternatives, aligns with our research objective: to empirically investigate the role of Generative AI in conceptual design rather than to evaluate the existing design-specific tools. Furthermore, as identified in the literature review, limited design-specific tools are designed to support the entire conceptual design process (Lee et al., 2024). Generative-purpose AI, therefore, is more suitable for achieving our research aim. The whole study procedure is shown in Figure 1. Primarily, we aimed to address three research questions through this experiment:

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press

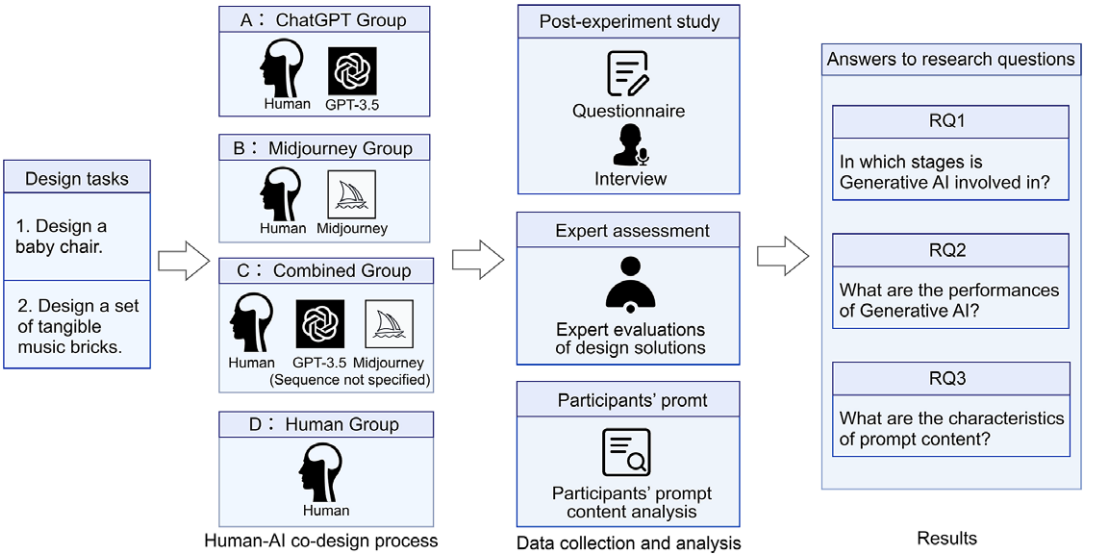


Figure 1. Representation of the experimental study procedure.

RQ1: In which stages is Generative AI involved in?

RQ2: What are the performances of Generative AI?

RQ3: What are the characteristics of prompt content?

3.1. Participants

We recruited participants through a university social networking site, with recruitment criteria requiring participants to be novice designers with <4 years of design learning experience. Additionally, it was required that participants have prior experience with ChatGPT and Midjourney. This can enable participants to focus more on design tasks rather than adapting to new Generative AI tools during the experiment. There were no restrictions for major background. After the screening process, a total of 20 participants (13 females, 7 males, aged 18–26, SD = 2.7) who met these criteria were selected. Each participant was paid $10 per hour, and the average time for completion was about 70 minutes.

3.2. Procedure

Participants were randomly divided into four groups, which are ChatGPT Group, Midjourney Group, Combined Group and Human Group. It is important to note that in the Combined Group, participants can use both ChatGPT and Midjourney freely for both tasks, without limitations on orders. Before the experiment started, we conducted a 20-minute training session that covered the basics of conceptual design procedures and how to utilize GPT-3.5 and Midjourney to generate conceptual designs. Then, the experimental procedure and two conceptual design tasks were introduced to each participant. Two distinct design tasks were selected to mitigate potential biases, such as participants' potential expertise in a single task-

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press

related area (Hu and Reid, 2018). The first task required participants to design a baby chair in 20 minutes and the second task involved designing tangible music bricks in 20 minutes. The selection of the two design tasks was based on two considerations: first, the design ideation of the task could be conveyed through shape and external structure instead of intricate details related to internal structure. Second, the conceptual design task should involve objects that participants are familiar with but not commonplace to ensure participants can complete the design tasks while having room for creative divergence. After introducing the experiment tasks and addressing the participants' questions about the procedure, the formal experiment began.

During the experiment, there are some specific requirements for each group:

- ChatGPT Group: Participants were asked to use ChatGPT to complete both of the design tasks.
- Midjourney Group: Participants were asked to use Midjourney to complete both of the design tasks.
- Combined Group: Participants were asked to use ChatGPT and Midjourney to complete both of the design tasks. The sequence of tool usage was not predetermined, allowing participants to choose the order based on their preferences.
- Human Group: Participants were asked to finish both of the tasks on their own (without the assist from Generative AI).

Finally, each participant was required to create an image for each task that illustrates the product's design features, accompanied by essential text descriptions to clarify the design.

After the two tasks, each participant was invited to fill out a questionnaire. The questionnaire encompasses a 7-point Likert scale across five criteria regarding the evaluation of Generative AI's performance (Participants in the Human Group were invited to evaluate their performance to serve as a baseline for comparing the performance of Generative AI). The questionnaire also explored the participants' reflections about which stages Generative AI helped with, and which stages are human-led (see details in Section 3.3.1). In the semi-structured interview, we discussed the questionnaire results with participants and their attitudes, evaluations and suggestions regarding the Generative AI-assisted human–AI collaboration. Each interview lasted around 30 minutes. All study procedures conformed to the Institutional Review Board (IRB) guidelines on human subject study.

3.3. Data collection and analysis

3.3.1. Participants' assessment of Generative AI tools performances

After completing the experimental tasks, a performance assessment questionnaire was distributed to each participant. In the questionnaire, participants needed to evaluate the performance of Generative AI tools which they used on six criteria: speed, subject, diversity, novelty, triggering more ideas and requirement satisfaction. The performance evaluation focused on the overall design process. These criteria were selected as they effectively reflect the impact of Generative AI in conceptual design. To be specific, the criteria for diversity and novelty were inspired by traditional conceptual design evaluations (Shah et al., 2003), while the other criteria (trigger more ideas, requirement satisfaction, speed and subject) were specifically formulated based on key factors in Generative AI-assisted

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press

Table 1. Participants' evaluation criteria

Criterion	Description
Speed	Measures how quickly the AI can complete conceptual design tasks, emphasizing efficiency in the design process.
Subject	Assesses whether the AI-generated outputs align with logistical requirements, ensuring they meet specific project needs.
Diversity	Evaluates the variety of ideas produced by the AI, reflecting its capability to explore a wide range of creative solutions.
Novelty	Rates the uniqueness and originality of the AI's outputs, crucial for innovation in design.
Trigger more ideas	Looks at the AI's ability to inspire further creativity among human collaborators, enhancing the ideation process.
Requirement satisfaction	Checks if the AI can accurately interpret prompts and produce results that fulfill user requirements, showing adaptability and responsiveness.

Notes: Participants in the Human Group were invited to evaluate their own performance to serve as a baseline.

conceptual design and related to the objectives of this study. Detailed explanations of the six criteria are described in Table 1. Additionally, for the Combined Group, the questions were designed to evaluate ChatGPT and Midjourney separately. For the Human Group, participants were asked to assess their own performance. This approach allowed us to collect firsthand feedback from human designers and gain insights into their modes of collaboration with AI.

3.3.2. Expert ratings

Five professional designers (three males, two females, aged 25–29), who have more than 5-year design experience, were recruited as experts to evaluate the conceptual design solutions created by the participants in the four groups. During the evaluation, the 40 design solutions were randomly displayed. For each solution, assessors were first introduced to which task (Task 1 or Task 2) the solution was from. Then, assessors were asked to assess the solutions using 7-Likert scale (1: The performance is really poor; 2: The performance is poor; 3: The performance is below average; 4: The performance is average; 5: The performance is above average; 6: The performance is very good; 7: The performance is perfect). The experts assessed the design solutions based on five primary criteria: (1) Novelty: whether the design introduces new ideas or approaches that are not commonly found in similar products; (2) Feasibility: whether the design can be realistically implemented; (3) Usability: whether potential users can easily and effectively use the product to achieve their goals; (4) Functional diversity: the range of functions that the design can perform; and (5) Cost: the overall expenses involved in manufacturing, distributing and maintaining the product over its lifespan (high-cost score means poor performance). The assessment process lasted around 30 minutes.

3.3.3. Generative AI's helpful stages in conceptual design

As this study aims to characterize human–AI collaboration in conceptual design at a stage level, we defined "actual-helping stages" as the stages where participants

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press

reported completing tasks with the assistance of Generative AI. To explore this, participants needed to fill the post-experiment questionnaire to report the stages in which Generative AI actually helped them. Additionally, participants needed to report which stages were led by humans, implying that human designers completed most of the work independently. These two questions were presented in the form of multiple-choice questions, allowing participants to select all the stages they felt were applicable. The questionnaire for the Combined Group is detailed in Appendix A as an example.

3.3.4. Participants' prompt

All text inputs by participants to communicate with Generative AI during their conceptual design process were collected. In total, we gathered 114 prompts across the ChatGPT Group, Midjourney Group and Combined Group for Tasks 1 and 2, averaging 3.8 prompts per participant per task. For the data analysis process, we first categorized participants' prompt into one of the four stages of conceptual design. Initially, a random sample of three participants' prompt from each group was selected, and two researchers independently categorized these samples to develop a preliminary understanding. After discussing their individual classifications and explanations, they reached a consensus, which facilitated the finalization of a comprehensive codebook, detailed in Appendix B. After establishing the codebook, the two researchers independently coded the prompts for the remaining two participants' prompt in each group, achieving an inter-rater reliability of $\kappa = 0.74$, indicating a strong agreement between the coders. Ultimately, in the four stages analyzed, there were 31, 14, 10 and 9 prompts identified with ChatGPT, and 13, 20, 5 and 12 prompts identified with Midjourney, respectively.

We then moved to the next phase of our analysis, which involved systematically summarizing the strategies ChatGPT and Midjourney assisted human designers with during each stage of the conceptual design process. Specifically, the same two researchers independently reviewed the categorized prompts to identify the assistance strategies provided by ChatGPT and Midjourney for each stage. Discussions were frequently made to resolve any discrepancies. Specifically, we applied the affinity diagramming method to aggregate and analyze the topics reflected in participants' prompt (Holtzblatt and Beyer, 1997). In this process, two researchers placed the original prompt contexts onto sticky notes, grouped them and iteratively labeled each group with descriptors to elucidate their shared themes. The summarized strategies, along with corresponding examples, are presented in Section 4.3.

3.3.5. Post-experiment interview

We conducted one-to-one interviews after the participants finished the two design tasks and questionnaire to gain deeper insights into how novice designers collaborate with Generative AI during the conceptual design process. The interview questions were tailored based on the participants' questionnaire responses and the design solutions they completed, focusing on the following aspects:

(1) Why and how did you use ChatGPT/Midjourney during the [specific design stage]?
(2) In the questionnaire, you rated [specific criterion] with [specific score]. Why did you give this rating?

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press

(3) In the conceptual design's human–AI collaboration, you mentioned that the [specific design stage] should be human-led. Why do you think so? Furthermore, you indicated that [specific design stage] requires collaboration between humans and Generative AI. Could you please explain this in detail?
(4) What are your other feelings and views about the collaboration between humans and Generative AI in conceptual design?

4. Results

In this section, the three research questions are systematically answered based on the analysis of the collected data. First, we answer the question of which stages that Generative AI is involved in in RQ1. Second, we explore the performances of Generative AI in the conceptual design process both from participants' perspective and expert ratings in RQ2. Third, in RQ3, we delve into the prompt analysis results from human designers.

4.1. RQ1: In which stages is Generative AI involved in?

We initially identified the stages where Generative AI assisted designers and those perceived as human-led. This analysis includes data from the ChatGPT Group, Midjourney Group and Combined Group. Figure 2 illustrated Generative AI's helping stages and human-led stages in conceptual design processes by two Sankey diagrams. The percentages represent the proportion of responses among the 15 participants. Figure 2 (a) demonstrates that Generative AI predominantly supported humans during the idea generation stage, problem definition stage and idea evolution stage. Respectively, 86.7%, 73.3% and 60% of participants recognized the assistance of Generative AI in these stages. This indicates that text-to-text and text-to-image Generative AI tools are particularly effective in

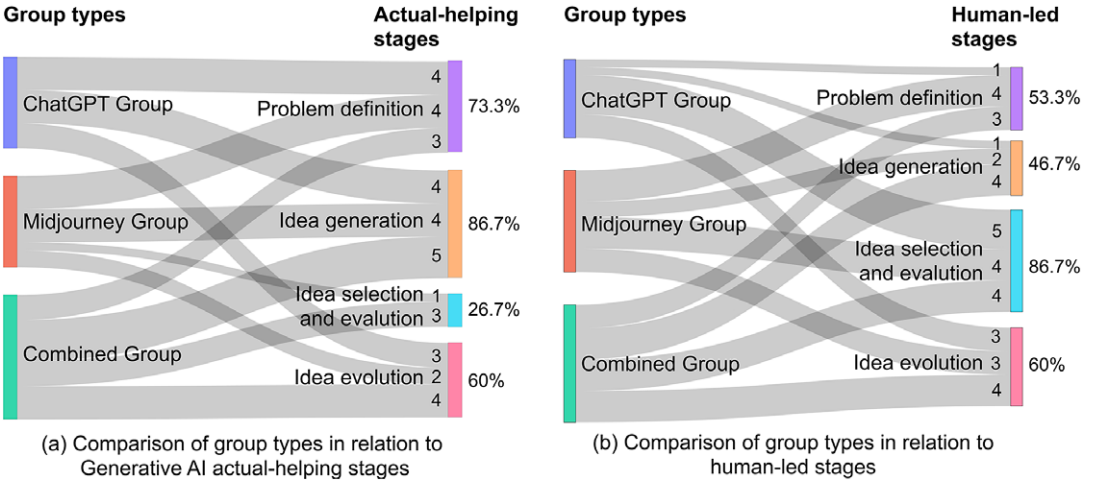


Figure 2. Horizontal Sankey diagrams representing (a) the comparison of group types in relation to Generative AI's helping stages and (b) the comparison of group types in relation to human-led stages. (Percentages in the figure represent the proportion of responses among the 15 participants in Generative AI-assisted groups.)

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press

initiating and nurturing early-stage design thinking, where conceptual blending and broad brainstorming are crucial (Wang et al., 2023). Figure 2 (b) reveals that the idea selection and evaluation stage (86.7%), as well as the idea evolution stage (60%), are predominantly perceived as human-led. This could imply that human judgment remains essential when it comes to evaluating and making final decisions on these ideas.

Overall, while Generative AI primarily supports the early stages of conceptual design, such as problem definition and idea generation, it still relies on human-led processes during the evaluation phase, highlighting the complementary roles of Generative AI and human expertise play in the design process.

4.2. RQ2: What are the performances of Generative AI?

This subsection synthesizes the assessments of Generative AI's performance by participants with expert evaluations of the final design solutions. By integrating these perspectives, we aim to provide a multidimensional understanding of how AI tools contribute to and influence the conceptual design process. As for data analysis, ANOVA analysis was employed when the data followed a normal distribution. For a non-normal distribution, the Kruskal–Wallis *H* test, a non-parametric test, was utilized to detect significant differences between the four groups.

Table 2 presents the assessments of five evaluation criteria by participants during the experimental process. Through the statistical results, some insights could be gleaned regarding model characteristics and human–AI collaboration. Notably, ChatGPT Group excelled in speed and requirement satisfaction. The superior speed performance can be attributed to ChatGPT's text-to-text model features, which allow for faster generation and real-time progress tracking. In contrast, Midjourney applies an iterative refinement process, starting with an initial visual pattern that progressively evolves into cleaner outputs through multiple enhancement steps. In addition, the requirement satisfaction scores were lower when participants used Midjourney, which means that instructions were not always effectively reflected in the final images produced. This reflects challenges in controlling Midjourney, as participants frequently reported that while they could

Table 2. Average scores and standard deviations of participants' evaluation of different Generative AI of each group in experimental design

Criterion	ChatGPT group	Midjourney group	Combined group	Human group*
Speed	**6.0** (0.71)	5.0 (1.00)	4.6 (1.14)	N/A
Subject	4.6 (1.14)	**5.6** (1.14)	5.2 (0.84)	4.8 (0.84)
Diversity	3.8 (1.48)	**5.2** (1.48)	4.6 (1.14)	3.4 (0.55)
Novelty	3.8 (0.89)	**5.4** (1.14)	4.0 (1.22)	3.6 (0.55)
Trigger more ideas	4.8 (1.30)	**5.6** (0.89)	4.6 (1.14)	4.6 (1.14)
Requirement satisfaction	**5.0** (1.87)	4.8 (0.84)	3.8 (0.84)	N/A

Notes:
1. * indicates our baseline condition.
2. **bold** indicates the best performance among four groups.
3. underline indicates performance worse than Human Group.

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press

manipulate shape design, the details did not align with their intentions (*P2-Midjourney*: "*When I input some commands, the generated images could generally shape the overall appearance, but the finer details did not align with my intended design specifications.*"). In contrast, Midjourney achieved the highest scores in terms of subject, diversity, novelty and triggering more ideas, likely benefiting from its visual representation which offers more direct stimuli. When comparing scores with and without Generative AI tools' help, ChatGPT Group achieved a lower score than the Human Group in subject, primarily because the Generative AI's outputs were probabilistic rather than fact-based (Brown et al., 2020).

The statistical assessment of expert ratings, detailed in Table 3, shows that the Midjourney Group achieved the highest overall score with a mean of 4.34 and a standard deviation of 0.75. Significant differences were observed in the metrics of novelty ($p < 0.01^{**}$), cost ($p < 0.01^{**}$) and overall performance ($p < 0.01^{**}$). Subsequent post hoc tests utilizing the Bonferroni correction revealed significant differences in novelty between the Human Group and the Midjourney Group ($p < 0.01^{**}$), Human Group and the Combined Group ($p < 0.01^{**}$). Additionally, the ChatGPT Group's mean score of 3.98 also surpassed that of the Human Group's 3.20, indicating that Generative AI tools could broaden the range of design options and introduce unique visual examples that enhance creativity. In addition, significant differences were found in the cost metric between the Human Group and both the ChatGPT Group ($p < 0.01^{**}$) and the Midjourney Group ($p < 0.01^{**}$). Notably, a higher cost score implies poorer performance, suggesting that the use of Generative AI may increase the complexity of design ideas, according to the definition of cost in Section 3.3.2. The same pattern was observed in overall scores, where both the ChatGPT Group and the Midjourney Group outperformed the Human Group, with $p = 0.02^{*}$ and $p = 0.01^{*}$ respectively. The Combined Group's mean score of 4.10 also surpassed the Human Group's 3.59. These results reflect that, compared with Human Group, human designers with Generative AI tools consistently achieved higher scores regarding expert ratings. The average Cohen's kappa among five assessors was 0.66 (with detailed results shown in Appendix C), which indicates an acceptable level of consistency.

Table 3. Expert rating results in combined two tasks

	Score (SD)					
Group	Novelty	Feasibility	Usability	Functional diversity	Cost	Overall
ChatGPT Group	3.98 (1.29)	4.56 (0.83)	**4.24** (1.12)	**3.88** (1.01)	4.52 (0.91)	4.24 (0.60)
Midjourney group	4.60 (1.23)	**4.62** (1.09)	4.10 (1.04)	3.60 (1.31)	4.76 (1.23)	**4.34** (0.75)
Combined group	**4.90** (1.31)	4.24 (0.89)	3.80 (0.78)	3.38 (0.96)	4.18 (1.25)	4.10 (0.66)
Human group	3.20 (1.17)	4.08 (0.77)	3.70 (0.92)	3.44 (1.08)	**3.52** (0.91)	3.59 (0.71)
P-value	0.000 **	0.084	0.228	0.422	0.001 **	0.005**

Notes:
1. ** denotes $p < 0.01$ and * denotes $p < 0.05$.
2. Bolded scores indicate the best performance among the four groups.

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press

4.3. RQ3: What are the characteristics of prompt content?

After completing the coding process described in Section 3.3.4, a summary of the strategies employed by human designers, alongside relevant examples, is presented in Appendix D due to space constraints. Subsequent sections will discuss these strategies as employed across the four stages of conceptual design: problem definition, idea generation, idea selection and evaluation and idea evolution.

4.3.1. Problem definition stage

For ChatGPT, the highest number of prompts was identified in this stage (31/64 = 48.4%), with five strategies summarized. First, it aided in **identifying the target audience** by providing demographic data, illustrated by a response detailing the age range for baby chair usage (*P5-ChatGPT Group*). Second, ChatGPT assisted in **user needs analysis**, as shown by an inquiry about parents' needs for baby seats (*P1-ChatGPT Group*), gathering insights that traditionally depend on extensive research (French et al., 1985). Third, participants utilized ChatGPT to offer insights into **existing products and their market status**, which streamlined the research phase by providing comprehensive and organized answers, exemplified by a prompt about current music visualization tools (*P1-Combined Group*). Fourth, **functionality considerations** were explored through inquiries about necessary features for a baby seat (*P5-ChatGPT Group*). Finally, participants also employed ChatGPT to investigate suitable **materials** that meet both the functional and aesthetic needs of the product (*P4-ChatGPT Group*).

For Midjourney, it enhanced the problem definition stage by **facilitating the exploration of various design intents**. It enabled designers to experiment with different aesthetic and functional styles visually, thus aiding the formulation of their own design concepts. A representative prompt is demonstrated by the P4-Midjourney Group "*baby chair*, *cute*, *bright colors*," which served as the initial input to explore a variety of design elements from an initial vague design direction.

4.3.2. Idea generation stage

ChatGPT facilitated idea generation in two distinct ways, differentiated by whether designers had initial design elements. These methods are identified as "**key design points synthesis**" and "**intuitive idea generation**." For instance, with specific design elements in mind, the P4-ChatGPT Group formulated a prompt to "*Design a baby chair that combines growth adaptability, non-toxic materials and music.*" Meanwhile, the tool could also generate original designs spontaneously without specific directions from designers, as seen in the P1-Combined Group's prompt: "*Design an innovative baby chair.*"

For Midjourney, the highest number of prompts was identified in this stage (20/50 = 40.0%), enhancing the idea generation stage in a relatively straightforward way by efficiently **transforms design ideas into visual representations**. This quick visualization saved much time and effort for manual sketching, which also facilitated the following idea selection and evaluation process.

4.3.3. Idea selection and evaluation stage

In the idea selection and evaluation stage, ChatGPT enhanced the design process by providing two essential types of support: **creativity evaluation** and **feasibility**

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press

assessment. Creativity evaluation primarily focuses on the novelty of design concepts. It involves ChatGPT aiding designers by highlighting innovative elements and suggesting areas for enhancement (*P4-Combined Group*). On the other hand, feasibility assessment concentrated on evaluating the practicality of the proposed concepts (*P1-ChatGPT Group*).

Meanwhile, Midjourney contributes by enabling the **creation of multiple design variants**. This feature facilitates the comparison and selection among different design concepts. For example, P3-Midjourney Group, who re-entered a prompt from an earlier step to generate more visualizations, explored various adaptations of a specific design idea. This process underscores Midjourney's capability to quickly adapt and visualize numerous iterations.

4.3.4. Idea evolution stage

In the idea evolution stage, designers primarily leveraged ChatGPT to enhance the design process in two ways: **refining design elements** and **facilitating concept iteration**. For the refinement of design elements, designers employed ChatGPT to improve and elaborate on the proposed solution's details. An example is the P3-ChatGPT Group's use of ChatGPT to refine a children's seat design by integrating more comfortable materials, as illustrated in the prompt: "*Refine the integration of fabric and Lego to optimize comfort and functionality in the children's seat design.*" In terms of concept iteration, designers revisited and revised their initial design directions. The revision process is exemplified by the P5-ChatGPT Group's request to "*Propose an alternative design for this baby rocking chair with modular components*," which shifted the focus from a standard design to one featuring modular components.

Meanwhile, Midjourney enhanced the idea evolution stage by **promoting design concept iteration** and **detailing visual enhancements**. Specifically, it facilitated rapid visualization and iteration of revised design concepts, enabling designers to swiftly explore and visualize modifications (*P2-Combined Group*). Additionally, it refined and detailed visual elements based on the same design theme, adding aesthetic intricacies that enriched the overall design (*P5-Midjourney Group*).

5. Discussion

5.1. The role of Generative AI in human–AI collaboration in conceptual design

Generative AI expands designers' solution exploration space and improves solution quality. Based on the results, all three Generative AI-assisted groups extensively utilized the provided Generative AI models during the idea-generation stage, as illustrated in Figure 2 (a). This extensive use is likely due to the contextual solution generation capabilities of Generative AI, a notable advantage highlighted in previous research (Wu et al., 2024; Weisz et al., 2024; Lee et al., 2024). From the perspective of designers' evaluation, the assistance of Generative AI facilitates the exploration of a broader solution space. Specifically, with the support of Generative AI, scores for diversity, novelty and the ability to trigger more ideas are all higher compared with those in the Human Group. From the perspective of design solution quality, experimental groups using Generative

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press

AI tools achieved higher mean scores across all five metrics examined in this study compared with the Human Group. Furthermore, overall scores for both the ChatGPT Group and Midjourney Group were found statistically significant differences compared with Human Group. This empirical evidence underscores the effectiveness of Generative AI in aiding novice designers during the conceptual design process. However, as this study aims to investigate "How Generative AI supports humans in conceptual design," our primary focus is on the role of Generative AI in "triggering more ideas." This focus may overlook the issue of design fixation potentially caused by Generative AI (Jansson and Smith, 1991; Wadinambiarachchi et al., 2024), leaving room for future research to explore this further.

With the assistance of Generative AI, human's idea selection and evaluation stage was further triggered. This finding stemmed from the post-interview data, where we asked participants to reflect on their design process from the perspective of conceptual design stages. When novice designers finish the design task on their own, the solution selection and evaluation stage may be overlooked (*P2-Human Group*, *P4-Human Group*). One possible explanation for this is that during the conceptual design process, designers often independently develop solutions starting from existing ideas (*P2-Human Group*: "*My strategy is that when I create this design, it was based on the existing possible problem with the baby and the stroller. This direct design process did not have an idea selection and evaluation process. It is a direct design and ignores the selection and evaluation processes.*"). In this context, they primarily engage in autonomous concept development and find it challenging to step outside their established cognitive frameworks to effectively evaluate and select among different solutions. However, the pattern changes when designers collaborate with Generative AI – the involvement in the solution selection and evaluation stage becomes more pronounced. This suggests that Generative AI may enhance the solution selection and evaluation stage by prompting designers to critically assess and justify the outputs it generates. This interaction may help break cognitive biases and encourage a more thorough evaluation process (*P2-Combined Group*: "*I reviewed everything ChatGPT and Midjourney generated. Some evident flaws would be found. Following that, I also got some new ideas about solving the problem.*").

Comparison between text-to-text and text-to-image models in initial conceptual design processes. Although the experimental results indicated that Generative AI primarily assisted in the problem definition and idea generation stages, text-to-text models and text-to-image models played distinct roles in these two phases. Specifically, in the problem definition stage, ChatGPT was able to outline key points of product design and provide suggestions for innovative designs (*P5-ChatGPT Group*: "*For the first task, I only have a general idea and did not know the specific details. So I asked GPT-3.5 to tell me what the needed functions should be. In the second task, I don't really know about how to design musical bricks, and I command GPT-3.5 to tell me what the design of musical bricks commonly encompasses and which aspects I could make innovations in*"), owing to the extensive knowledge base and the capacity for a certain level of reasoning. On the contrary, while Midjourney could offer help in the problem definition stage, it requires users to input solution-oriented prompts, which necessitates the user having a preliminary idea about the design solution, as P2-Midjourney Group noted: *P5-ChatGPT Group*: "*Because it (Midjourney)*

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press

relies on the initial keywords I provide. Without these keywords about design direction, it might deviate entirely from my intended idea." In essence, while Midjourney offers assistance in the problem definition stage, it is insufficient on its own.

Although ChatGPT excelled in helping designers analyze individual design elements in the problem definition stage, the integrated design solutions generated from the filtered design points may confuse novice designers at idea generation stage. Conversely, Midjourney's advantage in visualization saves designers time in expressing ideas related to shape, texture and color. For example, P1-ChatGPT Group inputted, *Provide design ideas based on the elements I give you*: "*a baby seat, appearance of Super Mario, blue and red as main colors. Integrate the above design points and make it more complete and detailed.*" However, ChatGPT's response remained in the form of key points (such as theme and color scheme, shape and features and fabric and materials). "*It could not provide me an overview of the design solution*," as expressed by P1-ChatGPT Group.

5.2. Implications for future conceptual design support under Generative AI's help

Workflow guidance and system integration should be carefully considered when combining text-to-text and text-to-image models. In our experiment, the combination of ChatGPT and Midjourney did not yield a synergistic effect, both reflected in the participants' assessment of Generative AI tools and expert ratings results. Interview results suggest a possible explanation, as P5-Combined Group noted, "*In the experiment, I primarily copied results from GPT-3.5 to Midjourney, but these models interpret my commands and produce results differently.*" This underscores how frustration could negatively impact the user experience with Generative AI, which might affect the outcome quality of human–AI co-creation solutions. Therefore, there is a necessity for methodologies and system designs that integrate the demands of various stages of conceptual design with the strengths of text-to-text and text-to-image models, respectively. Exploring better integration between these models could help leverage their combined potential.

Explore the effect of image stimuli on stimulating designers' inspiration. In previous research on Generative AI-enhanced conceptual design, the problem exploration stage was primarily enhanced by text-to-text models (Norheim et al., 2024), likely due to text being a fundamental mode of information expression. However, this study found that participants' feedback indicated the highest novelty scores were achieved by Midjourney, and in expert ratings, the Midjourney Group also obtained higher mean scores than the ChatGPT Group. Although the text-to-text model leverages a big knowledge base to compensate for designers' limitations in knowledge and experience, human designers may overlook potentially important details due to the extensive textual information. Therefore, future system development could consider aiding designers in integrating information output by text-to-text models with visual design elements, or exploring the potential of visual search (Son et al., 2024), which could help designers relate the LLMs' response to possible design solutions and enhance the role of image stimuli in inspiring designers' creativity.

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press

5.3. Limitations and future directions

In this study, we focused exclusively on two representative generative models: a text-to-text model (ChatGPT) and a text-to-image model (Midjourney). Our decision regarding the specific choice of input and output modality was twofold. First, the choice of text as the primary input modality was driven by its accessibility and familiarity, particularly for novice designers, facilitating easier expression of design intents. Second, for text and images as output modalities, they are regarded as the most commonly utilized data modalities in conceptual design, making them appropriate for our output modalities.

Regarding continuous technical enhancements, on the one hand, more modalities can be incorporated to facilitate more flexible and naturalistic communication into the human–AI collaboration process, such as integrating voice, video and sketches. By incorporating more modalities of Generative AI, researchers can more closely investigate the actual workflows of designers in experimental settings. On the other hand, with improvements in the generative models used in our research, such as GPT-4 and GPT-4o, researchers could explore two main directions in future work. One direction involves assessing their performances in processing multi-modal inputs. The other examines how these models perform in various types of design tasks, particularly those requiring more reasoning abilities, since previous research has revealed that as the complexity of tasks increases, the accuracy of the outputs generated by LLMs decreases (Khot et al., 2023). We believe these new avenues for subsequent empirical research could significantly contribute to the ongoing refinement and application of various Generative AI technologies across different design scenarios.

On the other hand, this study investigated the differences between groups assisted by Generative AI and those who completed tasks independently, which aims to uncover how these general-purpose Generative AI tools enhance designers' conceptual design processes compared with undertaking design tasks independently. Considering the specialized nature of design-specific tools, which are usually tailored to particular stages of conceptual design (Lee et al., 2024), future work could explore how workflow instructions and prompt engineering methods might affect the stages where Generative AI proves most beneficial.

For the choice of control group in this study, we selected Human Group for the purpose of comparing the differences and performances of human designers with and without the assistance of Generative AI. This approach helped us obtain some insightful findings and implications for future research. For example, with the assistance of Generative AI, the idea selection and evaluation stage was further triggered. Future work could include a comparative analysis with other traditional design support methods and tools, which would help provide a more comprehensive understanding of the value added by Generative AI.

Finally, our experiment revealed that participants in the Combined Group, despite not being restricted on the order of tool usage, consistently used ChatGPT first, followed by Midjourney. Investigating the impact of the sequence of tool usage on experimental outcomes could provide valuable insights. Additionally, expanding our study to include a broader range of participants, such as more experienced designers, could help validate and extend our findings across different levels of expertise.

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press

6. Conclusion

Our work aimed to investigate how Generative AI assists humans in the conceptual design process, especially for novice designers. Specifically, we conducted an experimental study involving 20 novice designers, assessing their performance with or without the help of text-to-text and text-to-image Generative AI models. The results revealed that Generative AI mainly assists humans in the initial stages of conceptual design, such as problem definition and concept generation, while the stages of idea selection and evaluation remain predominantly human-led. Despite the assistance of Generative AI, which improved participants' feedback and expert ratings, the combination of text-to-text and text-to-image models did not exhibit a synergistic effect. Based on the findings, we discuss the role of Generative AI in human–AI collaboration and compare the efficacy of different models in design assistance. Ultimately, we propose several implications for enhancing the effectiveness and user-friendliness of human–AI collaboration in conceptual design.

Acknowledgements

We thank all the participants for their time and the anonymous reviewers for their valuable comments. This research is supported by National Key R&D Program of China (2022YFB3303304).

References

Al'tshuller, G. S. 1999 *The Innovation Algorithm: TRIZ, Systematic Innovation and Technical Creativity*, Technical Innovation Center, Inc.

Brown, T., **Mann, B.**, **Ryder, N.**, **Subbiah, M.**, **Kaplan, J. D.**, **Dhariwal, P.**, **Neelakantan, A.**, **Shyam, P.**, **Sastry, G.**, **Askell, A.** *et al.* 2020 Language models are few-shot learners. *Advances in Neural Information Processing Systems* **33**, 1877–1901.

Cantamessa, M., **Montagna, F.**, **Altavilla, S.** & **Casagrande-Seretti, A**. 2020 Data-driven design: the new challenges of digitalization on product design and development'. *Design Science* **6**, e27.

Chen, L., **Cai, Z.**, **Cheang, W.**, **Sun, L.**, **Childs, P.** & **Zuo, H**. 2024a An llm-based concept generation method for solution-driven bio-inspired design. In *DRS2024: Design Research Society 2024 International Conference*, p. 908, Boston, USA.

Chen, L., **Jing, Q.**, **Tsang, Y.**, **Wang, Q.**, **Sun, L.** & **Luo, J**. 2024b Designfusion: Integrating generative models for conceptual design enrichment. *Journal of Mechanical Design* **146**(11).

Chen, L., **Song, Y.**, **Ding, S.**, **Sun, L.**, **Childs, P.** & **Zuo, H**. 2024c Triz-gpt: An llm-augmented method for problem-solving. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. **88407**, p. V006T06A010. American Society of Mechanical Engineers.

Chen, Q., **Shuai, W.**, **Zhang, J.**, **Sun, Z.** & **Cao, N**. 2024d Beyond numbers: Creating analogies to enhance data comprehension and communication with generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–14.

Choi, D., **Hong, S.**, **Park, J.**, **Chung, J. J. Y.** & **Kim, J**. 2024 Creativeconnect: Supporting reference recombination for graphic design ideation with generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–25.

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press

Deldin, J.-M. & **Schuknecht, M**. 2013 The asknature database: Enabling Solutions in Biomimetic Design. In *Biologically Inspired Design: Computational Methods and Tools*, pp. 17–27. Springer.

Design Council. 2019 The double diamond: 15 years on. *Design Council Journal.*

Eppinger, S. D. & **Ulrich, K**. 1995 *Product Design and Development*. McGraw-Hill

French, M. J., **Gravdahl, J.** & **French, M. J**. 1985 *Conceptual Design for Engineers*. Springer.

Gero, J. S. & **Kannengiesser, U**. 2014 The function-behaviour-structure ontology of design. In *An Anthology of Theories and Models of Design: Philosophy, Approaches and Empirical Explorations*, pp. 263–283. Springer.

Goodfellow, I., **Pouget-Abadie, J.**, **Mirza, M.**, **Xu, B.**, **Warde-Farley, D.**, **Ozair, S.**, **Courville, A.** & **Bengio, Y**. 2014 Generative adversarial nets. *Advances in Neural Information Processing Systems* **27**, 1–11.

Goodman-Deane, J., **Waller, S.**, **Bradley, M.**, **Yoxall, A.**, **Wiggins, D.** & **Clarkson, P**. 2016 Designing inclusive packaging. In *Integrating the Packaging and Product Experience in Food and Beverages*, pp. 37–57. Elsevier.

Guo, Z., **Song, M.**, **Fang, X.**, **Lin, C.**, **Zhang, H.**, **Li, X.** & **Wang, W**. 2024 Exploring synergies between aigc and triz in the optimisation of road cone design through integrated innovation methods. *Journal of Engineering Design*, **362**, 1–20.

Hatchuel, A. & **Weil, B**. 2009 Ck design theory: an advanced formulation. *Research in Engineering Design* **19**, 181–192.

Ho, J., **Jain, A.** & **Abbeel, P**. 2020, Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems* **33**, 6840–6851.

Holtzblatt, K. & **Beyer, H**. 1997 *Contextual Design: Defining Customer-Centered Systems*. Elsevier.

Hu, W.-L. & **Reid, T**. 2018 'The effects of designers' contextual experience on the ideation process and design outcomes. *Journal of Mechanical Design* **140**(10), 101101.

Jansson, D. G. & **Smith, S. M**. 1991 Design fixation. *Design Studies* **12**(1), 3–11.

Jasmine 2020 Conceptual Design: In 6 Steps. https://iit-techambit.in/conceptual-design-in-6-steps/ [accessed on 2024-08-10].

Jiang, S., **Hu, J.**, **Wood, K. L.** & **Luo, J**. 2022 Data-driven design-by-analogy: state-of-the-art and future directions. *Journal of Mechanical Design* **144**(2), 020801.

Karras, T., **Laine, S.** & **Aila, T**. 2019 A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Neural Information Processing Systems Foundation, Inc.; La Jolla, California, USA, pp. 4401–4410.

Kenton, J. D. M.-W. C. & **Toutanova, L. K**. 2019 Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of naacL-HLT, Vol.* **1**, p. 2. Minneapolis, Minnesota.

Khot, T., **Trivedi, H.**, **Finlayson, M.**, **Fu, Y.**, **Richardson, K.**, **Clark, P.** & **Sabharwal, A**. (2023) Decomposed prompting: A modular approach for solving complex tasks, In *Proceedings of the Eleventh International Conference on Learning Representations* (ICLR). Neural Information Processing Systems Foundation, Inc.; Kigali, Rwanda.

Kingma, D. P. & **Welling, M**. 2014 Auto-encoding variational bayes, In *Proceedings of the 2nd International Conference on Learning Representations (ICLR).* Neural Information Processing Systems Foundation, Inc.; Banff, Canada.

Kwon, E., **Rao, V.** & **Goucher-Lambert, K**. 2023 'Understanding inspiration: Insights into how designers discover inspirational stimuli using an AI-enabled platform. *Design Studies* **88**, 101202.

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press

Lee, S.-Y., **Law, M.** & **Hoffman, G.** 2024 When and how to use AI in the design process? implications for human–AI design collaboration. *International Journal of Human–Computer Interaction*, **413**, 1–16.

Liu, V., **Vermeulen, J.**, **Fitzmaurice, G.** & **Matejka, J**. 2023 3dall-e: Integrating text-to-image AI in 3d design workflows. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, MIT Press; Cambridge, Massachusetts, USA, pp. 1955–1977.

Luo, J., **Sarica, S.** & **Wood, K. L**. 2019 Computer-aided design ideation using innogps. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. **59186**, p. V02AT03A011. American Society of Mechanical Engineers.

Mahdavi Goloujeh, A., **Sullivan, A.** & **Magerko, B**. 2024 Is it AI or is it me? understanding users' prompt journey with text-to-image generative AI tools. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery (ACM); New York, NY, USA, pp. 1–13.

Myrup Andreasen, M., **Hansen, C.** & **Cash, P**. 2015 *Conceptual Design-Interpretations, Mindset and Models, Conceptual Design*. Springer; Cham, Switzerland.

Norheim, J. J., **Rebentisch, E.**, **Xiao, D.**, **Draeger, L.**, **Kerbrat, A.** & **de Weck, O. L**. 2024 Challenges in applying large language models to requirements engineering tasks. *Design Science* **10**, e16.

Radford, A., **Narasimhan, K.**, **Salimans, T.**, **Sutskever, I.** *et al*. 2018 *Improving Language Understanding by Generative Pre-Training*. OpenAI; San Francisco, California, USA.

Robles, G. C., **Negny, S.** & **Le Lann, J. M**. 2009 Case-based reasoning and triz: a coupling for innovative conception in chemical engineering. *Chemical Engineering and Processing: Process Intensification* **48**(1), 239–249.

Sarica, S. & **Luo, J**. 2024 The innovation paradox: concept space expansion with diminishing originality and the promise of creative artificial intelligence. *Design Science* **10**, e11.

Sarica, S., **Luo, J.** & **Wood, K. L**. 2020 Technet: Technology semantic network based on patent data. *Expert Systems with Applications* **142**, 112995.

Shah, J. J., **Smith, S. M.** & **Vargas-Hernandez, N**. 2003 Metrics for measuring ideation effectiveness. *Design Studies* **24**(2), 111–134.

Shahin, M., **Chen, F. F.** & **Hosseinzadeh, A**. 2024 Harnessing customized AI to create voice of customer via gpt3. 5. *Advanced Engineering Informatics* **61**, 102462.

Shi, F., **Chen, L.**, **Han, J.** & **Childs, P**. 2017 A data-driven text mining and semantic network analysis for design information retrieval. *Journal of Mechanical Design* **139**(11), 111402.

Son, K., **Choi, D.**, **Kim, T. S.**, **Kim, Y.-H.** & **Kim, J**. 2024 Genquery: Supporting expressive visual search with generative models In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery (ACM); New York, NY, USA, pp. 1–19.

Suh, S., **Chen, M.**, **Min, B.**, **Li, T. J.-J.** & **Xia, H**. 2024 Luminate: Structured generation and exploration of design space with large language models for human-ai co-creation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery (ACM); New York, NY, USA, pp. 1–26.

Tan, L. & **Luhrs, M.** 2024 Using generative AI midjourney to enhance divergent and convergent thinking in an architect's creative design process. *The Design Journal*, **274**, 1–23.

Vaswani, A., **Shazeer, N.**, **Parmar, N.**, **Uszkoreit, J.**, **Jones, L.**, **Gomez, A. N.**, **Kaiser, L.** & **Polosukhin, I**. 2017 Attention is all you need. *Advances in Neural Information Processing Systems* **30**, 20750–20762.

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press

Wadinambiarachchi, S., **Kelly, R. M.**, **Pareek, S.**, **Zhou, Q.** & **Velloso, E**. 2024 The effects of generative AI on design fixation and divergent thinking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery (ACM); New York, NY, USA, pp. 1–18.

Wang, S., **Petridis, S.**, **Kwon, T.**, **Ma, X.** & **Chilton, L. B**. 2023 Popblends: Strategies for conceptual blending with large language models In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery (ACM); New York, NY, USA, pp. 1–19.

Weisz, J. D., **He, J.**, **Muller, M.**, **Hoefer, G.**, **Miles, R.** & **Geyer, W**. 2024 Design principles for generative AI applications. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery (ACM); New York, NY, USA, pp. 1–22.

Wu, J., **Cai, Y.**, **Sun, T.**, **Ma, K.** & **Lu, C**. 2024 Integrating aigc with design: dependence, application, and evolution-a systematic literature review. *Journal of Engineering Design* 1–39.

Zhu, Q. & **Luo, J**. 2023 Generative transformers for design concept generation. *Journal of Computing and Information Science in Engineering* **23**(4), 041003.

Appendix A. Questionnaire example for participants in Combined Group

1. Which stages do you think ChatGPT had helped you? (You may select multiple options.)
   a) Problem definition
   b) Idea generation
   c) Idea selection and evaluation
   d) Idea evolution
2. Under ChatGPT's assistance, which stages do you think were human-led? (You may select multiple options.)
   a) Problem definition
   b) Idea generation
   c) Idea selection and evaluation
   d) Idea evolution
3. Which stages do you think Midjourney had helped you? (You may select multiple options.)
   a) Problem definition
   b) Idea generation
   c) Idea selection and evaluation
   d) Idea evolution
4. Under Midjourney's assistance, which stages do you think were human-led? (You may select multiple options.)
   a) Problem definition
   b) Idea generation
   c) Idea selection and evaluation
   d) Idea evolution

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press

Appendix B. The codebook for the stages classification of prompt

Table B1. Codebook for the prompt classification for four stages.

Stage	Definition
Problem definition	This stage involves the initial framing of the design challenge. It sets the foundation for design directions by gathering and synthesizing essential information to meet user and market needs.
Idea generation	This stage involves synthesizing the gathered information to formulate cohesive and innovative design ideas.
Idea selection and evaluation	This stage involves the assessment and refinement of generated ideas to select the most viable design concepts.
Idea evolution	In this stage, the selected design concepts are refined and detailed.

Appendix C. The Cohen's kappa results in expert ratings

Table C1. The Cohen's kappa results of expert ratings.

	Experts 1	Experts 2	Experts 3	Experts 4	Experts 5
Experts 1	/				
Experts 2	0.72	/			
Experts 3	0.78	0.92	/		
Experts 4	0.52	0.72	0.78	/	
Experts 5	0.37	0.58	0.64	0.55	/

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press

Appendix D. The strategies human designers employed when collaborating with Generative AI

Table D1. Human strategies for collaborating with Generative AI.

Stage	Tool	Strategy	Role	Example prompt
Problem definition	ChatGPT	Target audience definition	Helps identify the target user group for the product.	*"At what age do children usually need a crib?" (P5-ChatGPT Group)*
	ChatGPT	User needs analysis	Analyzes user requirements to guide design.	*"What are the needs of parents for baby seats?" (P1-ChatGPT Group)*
	ChatGPT	Related product research	Provides insights into current product types.	*"What music visualization tools are there now?" (P1-Combined Group)*
	ChatGPT	Functionality considerations	Explores functions and materials for design.	*"I need to design a baby seat. What are the commonly seen functions now?" (P5-ChatGPT Group)*
	ChatGPT	Material considerations	Investigates suitable materials.	*"What material should be used to make a child seat?" (P4-ChatGPT Group)*
	Midjourney	Design intent exploration	Explores aesthetic/ functional styles for design.	*"Baby chair, cute, bright colors." (P4-Midjourney Group)*
Idea generation	ChatGPT	Key design points synthesis	Integrates diverse elements into cohesive ideas.	*"Design a baby chair that combines score, mood, and Kandinsky into one toy." (P4-ChatGPT Group)*
	ChatGPT	Intuitive idea generation	Automatically generates design ideas.	*"Design an innovative baby chair." (P1-Combined Group)*
	Midjourney	Design idea visualization	Visualizes concepts rapidly.	*"Versatile and novel babychair…from 1 to 5 years old." (P5-Midjourney Group)*
Idea selection and evaluation	ChatGPT	Creativity evaluation	Evaluates creativity and suggests improvements.	*"How is the innovation in this baby chair…?" (P4-Combined Group)*
	ChatGPT	Feasibility assessment	Assesses practicality and constraints.	*"Can this baby rocking chair be adapted for compact urban homes?" (P1-ChatGPT Group)*

Continued

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press

Table D1. Continued

Stage	Tool	Strategy	Role	Example prompt
	Midjourney	Make design variants	Produces variants for comparison.	Retype the prompt from the previous step: *"adjustable, pushable, winter and summer, damping, baby seat, gray green, with toys, adjustable, with song rhyme pattern." (P3-Midjourney Group)*
Idea evolution	ChatGPT	Design elements refinement	Enhances key design components.	*"Refine the integration of fabric and Lego to optimize comfort and functionality in the children's seat design." (P3-ChatGPT Group)*
	ChatGPT	Concept iteration	Suggests new alternatives or adjustments.	*"Propose an alternative design for this baby rocking chair with modular components." (P5-ChatGPT Group)*
	Midjourney	Promote design iteration	Rapid visual feedback for idea evolution.	*Initial Prompt: "tangible music bricks, style like lego."* *Iterated Prompt: "tangible music bricks, color coding, instrument simulation." (P2-Combined Group)*
	Midjourney	Detailing and visual enhancement	Adds detail and improves visuals.	*Initial: "Tangible music bricks that children can use to explore musical sounds."* *Enhanced: "…adding vibrant color coding for different musical notes, textures that represent various musical instruments." (P5-Midjourney Group)*

https://doi.org/10.1017/dsj.2025.2 Published online by Cambridge University Press