Taylor & Francis
Taylor & Francis Group

Journal of Engineering Design

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/cjen20

A foundation model enhanced approach for generative design in combinational creativity

Liuqing Chen, Yuan Zhang, Ji Han, Lingyun Sun, Peter Childs & Boheng Wang

JOURNAL OF ENGINEERING DESIGN
https://doi.org/10.1080/09544828.2024.2356707

A foundation model enhanced approach for generative design in combinational creativity

Liuqing Chen[a], Yuan Zhang[a], Ji Han[b], Lingyun Sun[a], Peter Childs[c] and Boheng Wang[c]

[a]Department of Computer Science and Technology, Zhejiang University, Hangzhou, People's Republic of China; [b]Department of Innovation, Technology and Entrepreneurship, University of Exeter, Exeter, UK; [c]Dyson School of Design Engineering, Imperial College London, London, UK

ABSTRACT

In creativity theory, combining two unrelated concepts into a novel idea is a common means of enhancing creativity. Designers can integrate the Additive concept into the Base concept to inspire and facilitate creative tasks. However, conceiving high-quality combinational ideas poses a challenge that combinational creativity itself demands the consideration of conceptual reasoning and synthesis. We propose an AI foundation model enhanced approach for supporting combinational creativity. This approach derives combinational embodiments, and assists humans in verbalising and externalising combinational ideas. Our experimental study demonstrates that the generated combinational ideas by the approach obtained highest scores compared to those ideas generated without an AI foundation model or combinational strategy. We built a combinational creativity tool called CombinatorX based on this approach to generate ideas. In a study with the comparison of an existing combinational creativity tool and Internet search, we validated that our approach improves the effectiveness of combinational idea generation, enables a reduction in labour force, and facilitates the refinement of combinational ideation.

ARTICLE HISTORY
Received 31 December 2023
Accepted 13 May 2024

KEYWORDS
Combinational creativity; generative design; large language models; text-to-image models

1. Introduction

Artefacts with deliberate attention to design and aesthetics, depending on people's perspective and taste, tend to bring pleasure, such as some paintings, furniture, or architecture. The reason for this is that they are the creative products which are closely related to the creativity of the creators. Creativity is usually described as 'the ability to imagine or invent something new of value', and the process of transforming something novel and valuable to reality.

Boden (2004) identifies three approaches to achieve creativity in the human mind, including exploratory creativity, transformational creativity and combinational creativity, in which combinational creativity is regarded as the most accessible form of creativity. It is driven by combining unrelated or indirectly related concepts or ideas to produce new outcomes (Craft, Jeffrey, and Leibling 2001). To specify a simple and easy-to-implement

CONTACT Boheng Wang b.wang21@imperial.ac.uk Dyson School of Design Engineering, Imperial College London, UK

(a) A paper sculpture style teapot (b) A tree-shaped clothes hanger (c) A saddle-shaped chair

Figure 1. Innovative combination design works from Red Dot Design Award and IF Design Award, including (a) A paper sculpture style teapot (Ken Okuyama Design Co., Ltd. 2016), (b) A tree-shaped clothes hanger (Keith 2016) and (c) A saddle-shaped chair (Wei-Chieh and Hung-Hui 2023).

approach to combinational creativity, Han et al. (2018) propose a combinational creativity model of combining the Base and the embodiments from the Additive, where the Base is the main object of the combination, while the Additive is an additional supplement to the combination. A set of combinational creative design works from the Design Awards are shown in Figure 1, including a paper sculpture style teapot, a tree-shaped clothes hanger and a saddle-shaped chair. For instance, a paper sculpture style teapot can be described as a combination where the 'teapot' serves as the Base, and the texture of 'papercut' serves as the Additive.

In creative ideation, a good conceptual representation often offers more accuracy in describing the core of ideas (Lloyd-Cox et al. 2021), which is also valid in the context of combinational creativity. The form of representation using visual plus verbal form as a conceptual stimulus is often more effective in helping the designer transition to the subsequent design. However, verbalising combinational ideas and then visualising the ideas is challenging (Han et al. 2020). For people who are not adept at sketching or painting, creating combinational ideas is difficult and time-consuming (Bonnardel and Marmèche 2004). Even with leveraging robust computational capabilities of computers, simulating an entire combinational idea remains elusive. For example, the language models used to carry out concept verbalisation such as BERT often contain significant noise (Devlin et al. 2019), and generative models used to carry out concept visualisation such as GANs have in the past produced low-quality images (Goodfellow et al. 2014).

Recent advances in large language models (LLMs) seem to have brought a paradigm shift in creativity (White et al. 2023). In the field of natural language understanding, GPTs are general LLMs that have been trained on large-scale datasets, covering the major domains of human knowledge (Bian et al. 2023). By leveraging GPT's powerful inferring abilities, users can prompt initial natural language requirements to generate the outcomes that they are satisfied with. It can even undertake specific tasks such as code generation, logical problem-solving and summarisation, with supplying specific keywords and information through a few-shot learning (Zhang and Li 2021). In the field of image generation, the text-to-image (T2I) models based on Denoising Diffusion Probabilistic Models (DDPM) (Ho, Jain, and Abbeel 2020), have been trained on aligned pairwise text-image datasets, can interpret user inputs, identify analogous elements and features within extensive image datasets, and generate images meeting user specifications. However, LLMs and T2I models are not always useful or accurate, especially for intricate tasks such as generating combinational ideas. Users usually use trial and error to input the ideas to the model to see which output is a good one, but they do not know how to effectively manipulate the models and generate the results they are satisfied with. Thus, we pose a key research question: how to effectively leverage foundation models integrated with LLMs and T2I models to verbalise and visualise combinational ideas to support human creativity? To our best knowledge, there is no effective foundation model-based support tool for product design in the context of combinational creativity yet. The challenge lies in delineating tasks tailored to the respective strengths of LLMs and T2I models and devising generation strategies based on combinational creativity theory.

This study presents an AI foundation model enhanced approach that integrates GPT-4 and combinational strategy, for supporting combinational creativity. It diverges the embodiments associated with the Additive, assists designers in verbalising combinational ideas, and externalise the combined verbalisation into visual representations. As a foundation model, GPT-4 has been adopted as the combinational engine in the approach due to its representative and excellence in the area of LLMs and T2I models (Bang et al. 2023). The combinational strategy consists of two stages. Stage 1 includes a divergent thinking phase and a convergent thinking phase (Childs et al. 2022), which is delivered by LLM in GPT-4 and human respectively. LLM is used to simulates divergent thinking and generate embodiments from the initial Additive and verbalise the selected embodiment with the initial Base into a textual description to represent the combinational idea. Stage 2 is an externalisation phase that effectively visualises verbal combinational ideas to combinational images by T2I in GPT-4. The textual combinational description executes externalisation through the T2I function of GPT-4 for synthesising concept images containing both the base object and the Additive features. It is used to efficiently synthesise the morphology of the creative combinational idea and supplement the contextual elements of the product to further enrich the combination. Ultimately, the user can employ the generated verbal concepts as well as images as stimuli for subsequent design. Our approach aims to facilitate amateur designers and creative enthusiasts in accelerating conceptual designs and supporting creativity.

To summarise, this study makes the following contributions:

- The study presented an AI foundation model enhanced approach to support creative product design with combinational creativity for novices and amateur designers. The approach contains processes of divergent thinking, convergent thinking, verbalisation and externalisation, which are used to help users make effective combinations between different concepts and generate inspired combinational texts and novel combinational images.
- We explored how LLMs and T2I models can be integrated into combinational creativity and specified why and how LLM can be used to verbalise concepts and T2I can be used to externalise concepts.
- The study compared three combinational paths: Strategy Only, AI Only and Strategy + AI, to identify the effects of variables on combining concepts. While they are all effective, the integration path of the Strategy + AI produces more novel and original results.

- We emphasised the roles of the foundation model and human in the context of combinational creativity, and the implications of integrating generative AI with creative design.
- A case study showed that our approach achieved a lower perceived cognitive workload while achieving the level of automated tools (cutting labour by close to two times), compared with traditional Internet-based approach and computer-based automated product combination tool. We conclude our work with a discussion of the roles that people and AI may fill in interactive creativity support tools, and the unique value of integrating both.

2. Related work

2.1. Combinational creativity

Combinational creativity is also referred to as conceptual blending, emphasising generating novel ideas by exploring unfamiliar combinations of familiar concepts (Kaufman and Sternberg 2010). It can be realised by establishing explicit associations between concepts that originally had only subtle connections (Boden 2004). Combinational creativity-based creative generation holds significant implications in the field of creative design, with much of the research on creative synthesis focusing predominantly on derivative noun-noun combinations (Chen et al. 2019). Within these noun concepts, one is termed the Base, which serves as the primary or foundational concept in a creative combination, while the other is referred to as the Additive, acting as the supplementary concept in the formation of the combination (Wang et al. 2023c). For example, in the composite description of a 'lamp in the style of paper sculpture', the 'lamp' is the Base, and 'paper sculpture' is the Additive.

However, attempting to integrate disparate design elements in practical design scenarios is a notably challenging endeavour. Although some assisted creative design methods (Han et al. 2018; Shneiderman 2007) can significantly enhance design output quality and optimise the design experience, it is worth noting that these tools often encounter issues such as unsuccessful combinations, low-quality assembled results and outputs that still cannot be directly utilised. In recent years, researchers have proposed various methods and models to enact combinational creativity. For instance, Eppe et al. introduced a highly advanced computational framework (Eppe et al. 2018) to extend or generalise one or multiple inputs and search within these extensions to find connecting threads, thereby achieving creative combinations. Issa et al. constructed a knowledge base and proposed a programme for generating creative content based on combinational creativity (Issa, Alghanim, and Obeid 2019). However, both the generalisation of input data and the manual construction of knowledge bases and datasets entail high complexity and costs, since they are 'information-hungry' (Veale 2019).

2.2. AI foundation model: large language model and text-to-image model

An AI foundation model can be defined as a kind of large-scale generative model, which typically encompasses the functions of generating new content such as text, image, music (Huang and Guo 2019), audio (van den Oord et al. 2016), code (Weisz et al. 2021) and movement (Wallace et al. 2021). In this section, we primarily focus on language models and text-to-image models.

The evolution of language models implemented with deep neural networks can be traced back to Word2Vec, introduced in 2013 by Mikolov et al., which employs a shallow neural network to generate dense vector representations of words based on their contextual meanings (Mikolov et al. 2013). Further advancing the field, Vaswani et al. introduced the attention mechanism and crafted the Transformer architecture. Present-day state-of-the-art language models fall into two primary categories (Yang et al. 2023): those encoder-decoder, such as BERT (Devlin et al. 2019), RoBERTa (Liu et al. 2019b) and T5 (Raffel et al. 2020); and those decoder-only, including GPT (Brown et al. 2020; Radford 2019), OPT (Zhang et al. 2022), PaLM (Chowdhery et al. 2023) and BLOOM.

In the realm of text-to-image models, Variational Autoencoders (VAEs), conceived by Kingma and Welling, are set up to encode and decode input data (Kingma and Welling 2019). Venturing further into this domain, Aditya Ramesh et al. explored a direct approach to text-to-image synthesis leveraging autoregressive transformers, markedly amplifying the model's generalisability in zero-shot scenarios (Ramesh et al. 2021). At present, diffusion models have taken a leading position in the text-to-image generation domain. Models grounded in the DDPM framework, such as Stable Diffusion (Rombach et al. 2022), OpenAI's DALL·E 2 (Ramesh et al. 2022) and Google's Imagen (Saharia et al. 2022), have all demonstrated exceptional generative capabilities.

2.3. AI-Enhanced generative creativity design

Traditional online searches often fall short when confronted with simple common-sense queries or associative reasoning scenarios (Wang et al. 2023b). Therefore, introducing the robust search capabilities of AI often emerges as a more optimal strategy. This AI-enhanced approach offers a compelling foundation for the application of combinational creativity across various professional domains.

With the development of generative AI, an increasing number of scholars are investigating how artificial intelligence can support creativity and design in various domains such as creative image generation (Liu, Qiao, and Chilton 2022), conceptual association (Zuo et al. 2022), and architectural design (Tan and Luhrs 2024). For instance, Liu et al. (2019a) presented a method for Latent Space Cartography, which employs dimensionality reduction techniques to explore the latent design spaces within generative AI models. Wu and Li (2024) leveraged AI model to generate new images of knitted textile designs in fashion design. Since the intellectual property of AI-generated content belongs to the creator, the generated content can be used in real projects and creative sharing to support the creative process (*OpenAI* 2023).

Our literature review indicates that although AI foundation models can generate high-quality conceptual text or images, how to guide and control the AI-generated combinational ideas remains unexplored. While large foundation models have provided momentum in the field of generative design, they currently lack the capability to create novel and ground-breaking design concepts (Bender et al. 2021). This limitation primarily stems from their training on existing data and knowledge, which may contain biases or outdated information. Furthermore, although large models can generate content that is grammatically correct and contextually appropriate, they may fall short in deeply understanding complex design requirements and constraints, such as those in combinational creative design, and struggle to adapt flexibly to evolving design contexts (Marcus 2020).

3. Methodology

3.1. Overview

As depicted in Figure 2, the foundation model enhanced approach represents a two-stage automated conceptual design pipeline that is predicated on both LLM and T2I. This innovative method synergistically merges disparate concepts and elements, thereby engendering novel and coherent creative ideas. In stage 1, we engage in divergent thinking around the Additive to obtain corresponding embodiments, among which six specific features serve as directions for divergent thinking. These embodiments are inferred by the LLM through a well-defined prompt template. Generated embodiments subsequently undergo a process of meticulous selection and convergence. The embodiments are then amalgamated with the Base input to formulate composite textual combinational ideas, which are fed back into the LLM. Once these textual combination ideas are obtained, the externalisation is executed in stage 2, where images are strategically generated through T2I. To facilitate the T2I in generating high-quality composite images, we have engineered both an appearance generator and a scene element generator. The appearance generator is designed to enable efficient synthesis of visual concepts that encompass both the primary features of the Base and the Additive. On the other hand, the scene element generator employs visual metaphor theory to create metaphorical visual elements based on the product's functionalities and usage scenarios, thereby augmenting the richness of the product's visual conceptualisation. The process of appearance and element generation is called 'visual templates'.

3.2. Stage 1: find the additive embodiments and form textual combinational ideas

3.2.1. LLM inputs

As the Base is the main feature of the idea, the degree of creativity in a combinational idea largely depends on how to introduce a novel Additive. Generally, the combinational design process carries a certain level of abstraction, and there is a lack of explicit connections and concrete expressions between different concepts. An effective way to reduce the combinational complexity is to ideate the related associations from the Additive before the combination. Therefore, at the start of the process, we positioned the Additive as the input of the system to help designers jump out of the box to achieve cross-contextual thinking. By utilising the immense knowledge and reasoning capabilities of LLM, the association search for the Additive can be realised.

3.2.2. Divergent: embodiments search

To enable the combination of the Base and the Additive, and display the combinational pathway between them, the primary task in stage 1 is to identify and diverge the associations of the Additive, which lay the foundation for the subsequent combination with the Base. However, the Additive has multiple features, making the intuitive filtering process quite challenging. Without a thorough evaluation and judgment of the quality of these features, it is difficult to determine which features truly contribute to the innovation and

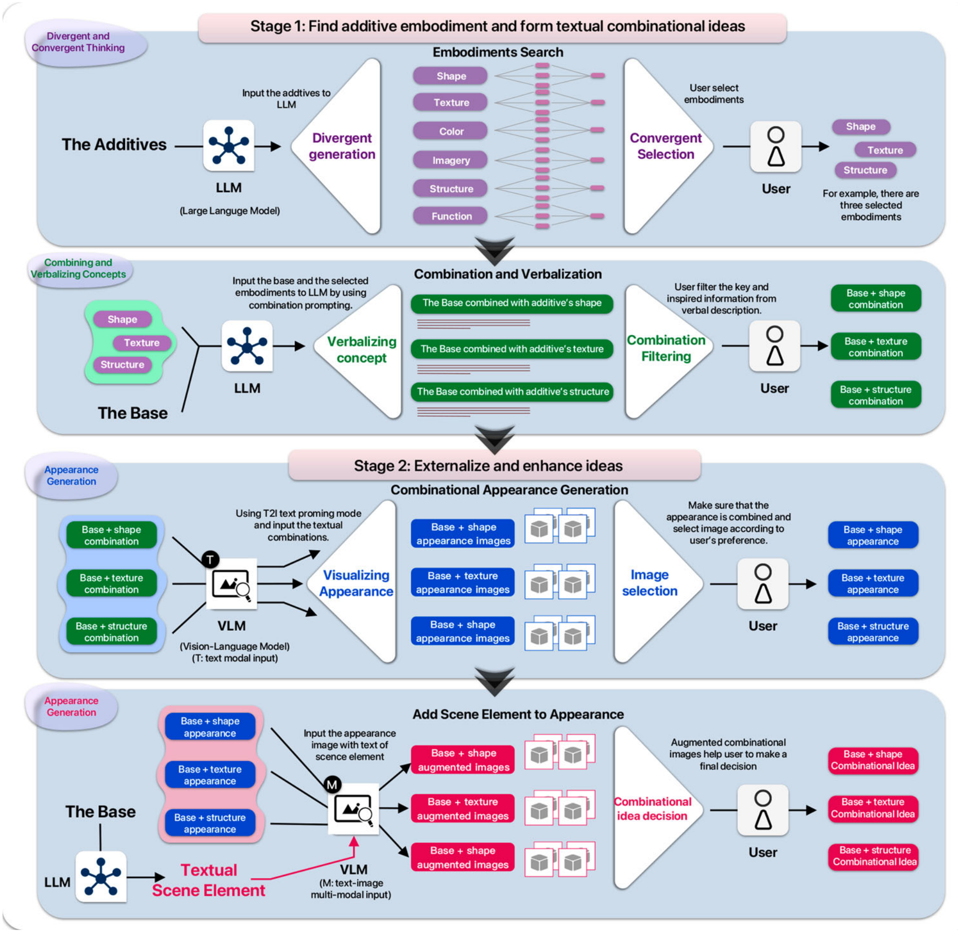


Figure 2. The diagram of the approach for Stage 1 and Stage 2, where we explore four different phases to visualise combinational ideas: Embodiments Search, Combination and Verbalisation, Combinational Appearance Generation and Add Scene Elements to Appearance.

utility of the combinational design. To address this issue, we drew inspiration from a classic analogy in design methodology – biomimetic design (Bar-Cohen 2006). Based on this, we categorised and indexed these features, ensuring that each of the Additive diverges from a fixed set of six features which are shape, texture, colour, imagery, structure and function. We call these features as 'embodiments'. Subsequently, a comprehensive evaluation and filtering process was conducted within each embodiment. The different embodiments were composed into 6 combinational options with the Base through LLM, and we call those options 'combinational templates'.

The embodiments can be considered as the key elements for describing or defining objects and concepts. For example, when we describe an object, we usually first consider its shape, texture and colour. Each of these features carries a unique dimension of information that ensures a comprehensive description. Elements such as shape and texture primarily reveal the visual characteristics of an object, while structure and function focus on its practical applications and value. Although there are various other features to be concerned with, the rationale for focusing on these main six features is to simplify workflow, increase efficiency and avoid unnecessary complexity in the design process.

To delve deeper into the latent semantic features of the Additive, we employ an approach that analyses the key attributes, thereby aiding us in identifying the corresponding embodiments. Taking 'hammer' as an example, a scrutiny of its functional features readily associates it with related terms like 'striking' and 'tool'. Moreover, chain reasoning techniques like mental linking fully harness the robust capabilities of the LLM (Diao et al. 2023). We provide detailed templates and full examples of prompts in Appendix A.

3.2.3. Convergent: filtering embodiments by human selection

For the convergent stage, we adopted a human-centred strategy and enhanced the user experience by letting them lead the decision-making process over the embodiments. To alleviate the decision-making burden, we impose no limits on the number of embodiments the user selected and allow for personal preference choices. If the user expresses dissatisfaction with the initial results, they have the option to request the LLM to regenerate output, offer feedback or specify requirements for a more desirable output. Additionally, users could directly edit the embodiments. This process may take several iterations until the outputs align with the user's expectations, after which the system seamlessly transitions to the subsequent stage.

3.2.4. Verbalising combinational ideas

The filtered embodiments and the Base as new inputs are collectively fed into the LLM to combine into textual concepts. By explicitly specifying the Base and the Additive, we provide the LLM with a well-defined verbal quest, which can effectively minimise ambiguity and directional confusion in the design. Within this prompt, we also include an encouraging key prompt: 'Please fully exercise your imagination'. This aims to let LLM simulate the process of imaginative thinking to elicit creative responses that transcend traditional designs. We then introduced the filtered embodiments to allow designers to focus on specific design details – such as the structural features in this example – to ensure they conveyed emotional connections. This added complexity refines the design task and provides a clearer design trajectory for the LLM.

After obtaining the integrated design concepts, these concepts must be further filtered or optimised based on their quality. First, a successful combinational design should follow its core principle, which means that the final design product should reflect the Base, while seamlessly integrating the functions of the Additive. If the synthesis falls short of expectations, potential manifestations may include a final product that is incongruent with the Base or does not adequately reflect the properties of the Additive. For those sub-optimal design outcomes, the designer has the option to regenerate or fine-tune the details. It is worth noting that although the Base can theoretically be combined with multiple Additives, this study focuses on scenarios involving a single Additive. Correspondingly, each embodiment represents a specific way of combination.

3.3. Stage 2: externalise and enhance concept

3.3.1. T2i inputs

In the context of T2I, achieving precise and unambiguous image output requires the input text to be clear, specific and devoid of vagueness or excessive abstraction.

Keyword-formatted inputs are considered optimal within the T2I due to their explicitness and conciseness. In this study, the combined sentences generated by LLM contain noise from other information. A T2I may not be able to generate images accurately. To address this issue, we employ key phrase extraction techniques to convert the textual design concepts generated in stage 1 into a list of key phrases, which are used as inputs. To minimise the noise of introducing scene elements to T2I and keep the appearance of the combinational idea unchanged, the scene elements generator is a multimodal input. It will receive an image generated from the appearance generator and a scene element obtained from the Base and then generate the enhanced combinational image.

3.3.2. Appearance generator

To effectively visualise visual design of the combinational concepts, it is crucial to establish the Base as the main component of the integrated product. This ensures that the primary function of the combinational design remains aligned with the core functions of the Base in practical applications. For example, if the Base is a cup and the Additive is a clock, this implies that the primary use of the resulting composite product is as a cup.

Generative large-scale models are not capable of deep understanding grammar and sentence structure in the same way that humans do. Their performance largely depends on textual descriptions, commonly referred to as 'prompts', and their associated keywords. Effectively organising prompts and keywords can enhance the visual appeal of images generated based on the same description (Pavlichenko and Ustalov 2023). Accordingly, we extract a list of keywords from the textual combination concepts generated in the first phase using LLM, to better serve as input for the image generator. For instance, we subjected our design idea of a hanging cup to keyword extraction and obtained an optimised prompt, which reads, 'Pendulum cup, traditional bell, elegant and charming design, side of the cup, waist-shaped, cylindrical, miniature clock, made of metal, hanging device, metal chain, clean white background, 8k'. To ensure that each output image maintains its original quality and consistently presents a clean background, we included 'clean background' as a core descriptor in each prompt.

3.3.3. Scene elements generator

The Scene Element Generator is predicated on visual metaphors, capable of generating metaphorical visual elements based on the product's functionality and usage context as supplementary content to the product's visual concept. We employ LLM to assist in exploring the metaphor-based visual representations related to the Base. It analyses the core functional meaning and related keywords of the Base metaphors in a comprehensive and detailed way. For instance, the primary function of a hair dryer is to rapidly dry hair or other objects through airflow, with 'airflow' being the keyword. Guided by this keyword, LLM can generate visual features that intuitively convey the corresponding metaphorical function, such as using dynamic lines to represent airflow.

To maximise the retention of details in the high-quality images generated by the appearance generator, we adopt a text-image multimodal input, which consists of a metaphorical text derived from the Base and an image generated from the appearance generator. Multimodal input aids the model in making more precise and efficient interpretations when faced with ambiguous or unclear directives. This effectively mitigates the risk of the model misidentifying visual elements intended for conveying metaphors as product

Figure 3. Using a hair dryer (a) as the foundational element, the scene element generator employs dynamic lines to simulate the visual effect of air flow, ultimately yielding a harmonious and aesthetically pleasing product design (b, c).

design elements, thereby ensuring that the generated product images are both accurate and closely aligned with user needs.

As illustrated in Figure 3(a), we showcase an image of a hairdryer set against a clean background. This image, along with textual visual elements – such as 'wind' and 'flowing lines' – comprises the multimodal input for the T2I's processing and interpretation. Following comprehensive analysis and rendering by the model, the generated product images are presented in Figure 3(b,c).

4. Method evaluation

In this section, we evaluated the method, which involves the capabilities of divergent reasoning, combinational ideas verbalisation, and concepts externalisation. Since the synthesis process of the approach is involved in both stage 1 and stage 2, we set a combinational strategy from stage 1 to stage 2 with AI models including LLM and T2I as two variables and evaluated the output generated from both LLM and T2I. As shown in Figure 5, both combinational strategy and visual strategy can either integrate AI or not. In our method, this is the integrated AI case. Therefore, we established three pathways to examine reasoning and visualisation capabilities, including Strategy Only, AI Only and Strategy + AI to answer one of our research questions – which variables can enhance the quality of combinational creativity?

4.1. AI foundation model configuration and case selection

We selected ChatGPT based on GPT-4 as the foundation model for our evaluation tasks due to its status as the most representative and widely used large model currently available and can play the role of both LLM and T2I. ChatGPT's accessible API makes a quick instruction of LLM deployment for those designers who unfamiliar with AI development.

For a creative product combination, the Base is a concrete object that forms the main function of combinational design. In contrast, the Additive can be either abstract elements or concrete objects. Thus, we selected two representative cases of product design

(a) LOVE Chair (b) Cup for Liquor

Figure 4. Combinational creative designs collected in the Red Rot Design Award and IF Design Award, including (a) LOVE Chair (Eugeni 2020), and (b) Cup for Liquor (Koizumi and Yuichiro 2021).

containing combinational creativity from the IF Design Award for technical assessment. The first case, as shown in Figure 4(a), is a combinational creative design based on an abstract concept, inspired by a multi-dimensional interpretation of love. In this design, the elements we artificially extracted as the Base and the Additive are chair and love, respectively. The second case, depicted in Figure 4(b), is a combinational creative design based on a concrete concept: a cup that produces pleasing chimes when raised for a toast. In this case, the Base and the Additive are cup and bell, respectively. Both cases offer rich potential for creative divergence and pose certain design challenges.

4.2. Experimental controls

In our experiment, participants were required to engage in combinational creative design based on provided cases and tools. We recruited 30 novice designers with comparable design capabilities to participate in this design challenge. Participants were evenly distributed into three groups, where each group had 10 individuals. All participants had an undergraduate educational background in design, spanning 3–4 years. To control for experimental variables and minimise biases in the design process, each participant received training before the commencement of the experiment on how to effectively utilise the methodology attributed to their designated group. The entire design generation process was constrained to a 40-min timeframe, and participants were expressly prohibited from using any other software or search engines.

The experimental setting is shown in Figure 5. Five members work on case 1 and five members work on case 2 in each group. The first group adopted the Strategy Only approach, where participants generated design solutions based exclusively on pre-supplied combinational templates in stage 1. Once they obtained verbal ideas. They were required to use non-AI tools such as Adobe Photoshop to visualise the ideas by using visual templates in stage 2. The second group, denoted as AI Only, required designers to generate verbal design ideas autonomously without the combinational templates in stage 1. Then they were required to freely use ChatGPT to visualise the idea without visual template guidance in stage 2. The third group, termed Strategy + AI, utilised the AI-enhanced

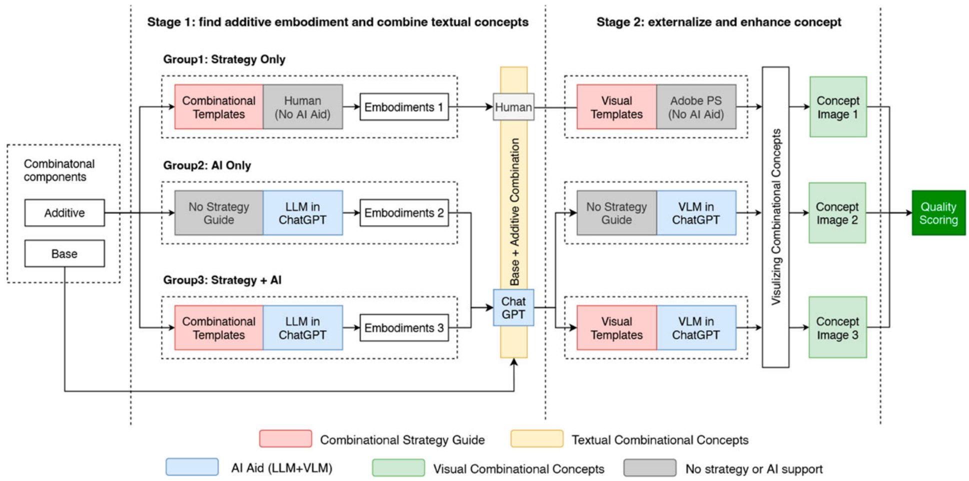


Figure 5. An experiment for technical evaluation, where participants were divided into three groups: Strategy Only, AI Only and Strategy + AI.

method, combining both combinational templates and visual templates, guiding ChatGPT for creative product design.

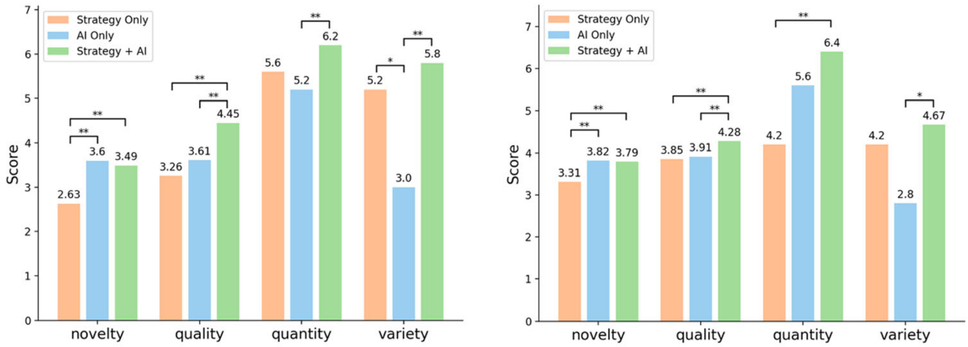
4.3. *Creativity evaluation*

To accurately assess design concepts, 6 experts with over 5 years of design experience were employed as the assessors. All ideas were evaluated under a unified set of guidelines and inter-rater agreement on a 7-point Likert scale. Four criteria were employed for the evaluation: quantity, novelty, quality and variety, following the assessment metrics proposed by Shah, Smith, and Vargas-Hernandez (2003). Quantity was gauged by counting the total number of design ideas generated by each individual, while variety was quantified by the types of design concepts, which were classified based on various combinations of the Base and the Additive.

To quantify the level of agreement between the assessors, a Fleiss Kappa test was conducted. Test results indicate that in stage 1, for cases 1 and 2, the Fleiss Kappa coefficients for novelty, quality, quantity and variety were respectively 0.693, 0.731, 1 and 1, and 0.669, 0.651, 1 and 1. In stage 2, the coefficients were 0.685, 0.772, 1 and 1, and 0.660, 0.687, 1 and 1, respectively. This reveals almost perfect agreement among the assessors in terms of quantity and variety, and substantial agreement for novelty and quality. Hence, it is valid and meaningful to use the average scores from the 6 evaluators as the final assessment outcome. Based on the psychometric evaluation, the mean scores of quantity, quality and novelty at the individual level for stage 1 and stage 2 were calculated and presented in Figures 6 and 7 respectively.

4.3.1. *Evaluation of stage 1*

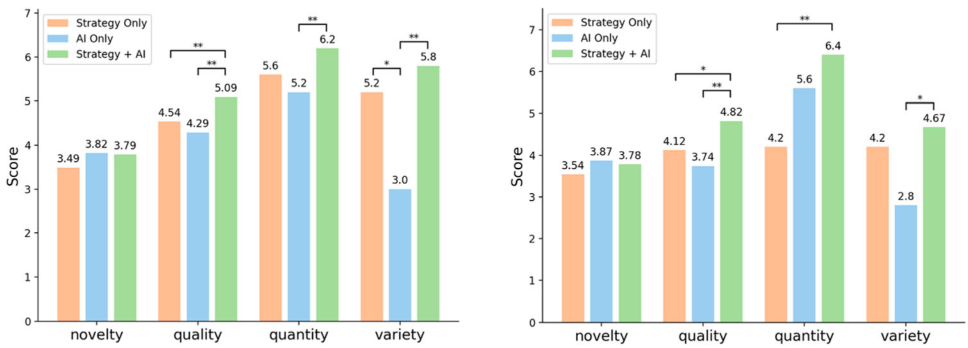
As depicted in Figure 9, in both case 1 and case 2, the novelty scores of the Strategy + AI and AI Only groups showed differences compared to those of the Strategy Only participants. Regarding the quality scores, there were significant disparities between



(a) Stage 1, Case 1: chair + love

(b) Stage 1, Case 2: cup + bell

Figure 6. Expert evaluation results across three design strategies of Case 1 and Case 2 in Stage 1.



(a) Stage 2, Case 1: chair + love

(b) Stage 2, Case 2: cup + bell

Figure 7. Expert evaluation results across three design strategies of Case 1 and Case 2 for Stage 2.

Strategy + AI and both AI Only and Strategy Only participants. On an individual level, the output amount between the Strategy + AI and Strategy Only displays a pronounced difference. For the variety, participants using Strategy + AI in case 1 presented an average of 5.8 idea categories, compared to 5.2 for those relying on Strategy Only, while AI Only participants showcased an average of 3.0 categories. Notably, both Strategy + AI and Strategy Only exhibited significant differences when compared to the AI Only participants.

4.3.2. Evaluation of stage 2

As depicted in Figure 7, across both case 1 and case 2, the novelty scores of the three methods showed no significant differences. In both cases, the quality scores of Strategy + AI demonstrated significant differences when compared to those of Strategy Only and AI Only. Specifically, in case 1, the quality scores for Strategy Only, AI Only and Strategy + AI were 4.54, 4.29 and 5.09 respectively. In case 2, the scores were 4.12, 3.74 and 4.82 correspondingly. The quantity and variety of the results remained consistent with those of stage 1.

4.3.3. User interview

Upon the conclusion of the experiment, we conducted user interviews by selecting one participant from each group for every case to address the following questions regarding their experience during the experiment. Q1 and Q2 are to investigate the verbal process experience in stage 1. Q4 and Q5 are to investigate the visual process experience in stage 2. Q3 and Q6 are to investigate whether the two stages can successfully implement the combination separately.

Q1) 'Did you find it challenging to search for embodiments? If so, why?'

Q2) 'Did you encounter any difficulty in combining the embodiments with the Base?'

Q3) 'Did the resulting combination meet your expectations? Does the textual description comprehensively capture the features of both the Base and the Additive?'

Q4) 'Did you find it challenging to execute the visualisation of combined images? Why or why not?'

Q5) 'How would you rate the quality of the visualised image combination? Does it align with your ideal representation?'

Q6) 'Did the image consistently incorporate features from both the Base and the Additive?'

In stage 1, participants relying solely on the LLM conveyed difficulties in the search and design combination of embodiments. The information provided by the LLM was primarily in textual form, which wasn't always intuitive, and at times posed comprehension challenges. One user commented, 'The LLM does not always go into detail when dealing with questions, such as there are mutually exclusive options in the answers given'. Additionally, feedback from another user indicated, 'It's difficult to quickly come up with an effective way to ask questions. Even with repeated inquiries, the LLM's responses tend to be basic and superficial'. In contrast, designers employing both strategy and AI remarked, 'Searching and combining embodiments based on provided prompts with LLM is rather effortless. Although LLM's outputs might occasionally seem commonplace, it consistently offers a plethora of clear design suggestions, making it simpler to filter superior designs'. Notably, all participants interviewed during the later stages expressed satisfaction with the merged design concept, highlighting its successful integration of both the Base and the Additive characteristics.

Moreover, participants depending solely on strategy experienced hurdles in divergent design concerning specific features, particularly in the realms of function and structure. This was especially evident in case 1, where the abstract nature of the Additive concept intensified the challenge. Consequently, in Chapter 5, we adopted this case as the subject of our case study.

In stage 2, when confronted with the complexities of visualisation, AI-assisted groups observed, 'With the aid of AI, the visualisation process has become relatively straightforward, even effortless'. In contrast, those adhering strictly to strategy noted, 'Executing multiple final product drawings manually within a constrained timeframe is indeed challenging'. However, AI-exclusive groups reported a more considerable divergence from their initial expectations. One participant stated, 'Even though the textual descriptions might be relatively comprehensive, the images produced by feeding the text into T2I does not always align with the initial vision. Constant prompt optimisation is required, which isn't always

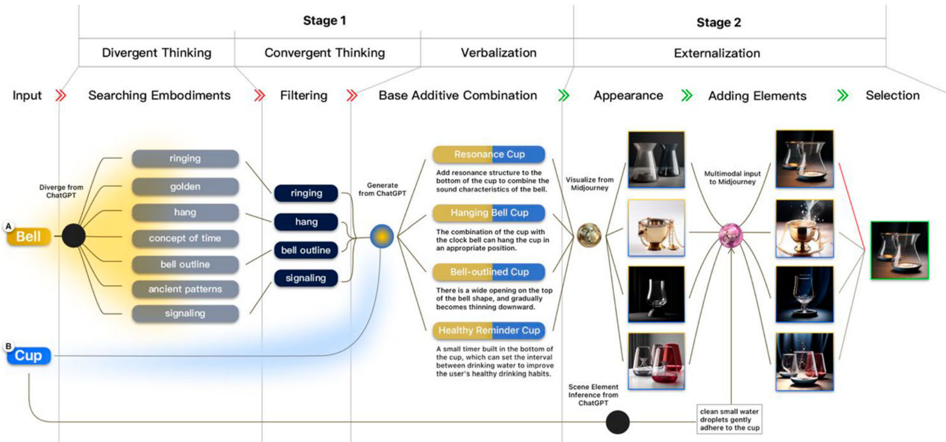


Figure 8. An example of the CombinatorX system generates combinational ideas and images for the inputs of the cup and bell. Initially, LLM diverges from the concept of a bell to produce various embodiments. Subsequently, designers converge and filter out select high-quality embodiments to combine with the cup, and the outcomes are visualised. To enhance the expressiveness of the product images, LLM creates visual elements that convey the metaphorical functions of the cup.

straightforward to grasp'. Simultaneously, while the Strategy + AI groups acknowledged that some design details could not be entirely captured in the visualisations, none explicitly stated a significant deviation from their anticipations. Nevertheless, all participants unanimously agreed that the resulting images encompassed characteristics of both the Base and the Additive.

5. Combinatorx and case study

5.1. Combinatorx

We implemented CombinatorX, a combinational creativity tool based on the foundation model enhanced approach. Figure 8 shows an example of the process of using this tool for a combination of cup with bell.

To determine whether and how CombinatorX facilitates the creation of combinational creative concepts, we conducted a within-subject study to compare it with two baselines. One is traditional internet-searching-based design method, which allows users to use any search engine to gain fragments of ideas and to illustrate a combinational idea in sketching. Another one is called 'Combinator', which is an automated creativity tool for combinational design (Han et al. 2018). It represents a typical computer-based automated approach and is inspired by the analogy-based human brain's ideation. Specifically, our evaluation aimed to identify whether users were able to: (1) create higher-quality combinational product design and (2) experience a reduced ideation difficulty for combining two concepts by using CombinatorX. In addition, the combined ideas by our approach were also validated through a modelling task. Participants with the approach based on the foundation model were asked to design the 3D model as the design implementation of combinational idea by referring the generated idea by CombinatorX. This aims to identify that the design generated by the proposed approach is worthwhile.

5.2. Participants

We recruited six novice designers with similar levels of product design who have 3–4 years of undergraduate design education background. To minimise the bias of the study, participants were required to demonstrate proficiency with the design tools they were asked to use.

5.3. Procedure

The participants were tasked to execute combinational designs using assigned tools to create the combination of 'chair + love'. Of the six participants, two used CombinatorX, another two used Combinator, and the remaining two relied on search engines and sketching for creative combinations.

At the beginning of the experiment, participants were introduced to the fundamental principles of combinational creativity. They were asked to generate as many number of concepts as possible within a span of 30 min. Upon completion, they filled out the NASA-TLX questionnaire. The NASA-TLX evaluates their perceived cognitive workload, which includes mental demand, physical demand, temporal demand, performance, effort and frustration. In a subsequent 10-min semi-structured interview, questions centred around their experience and their perspectives on the system's outputs were posed. Considering the varying levels of quality of one person's designs, they were asked to select and present the best design concept they thought and its corresponding image.

In order to determine whether the generated concepts using our methodology have benefits for subsequent prototyping as well as engineering implementation and meet the proposed design requirements, the participants using CombinatorX were asked to take part in an additional modelling task after obtaining the generated concepts. They were required to design a 3D model based on the text and corresponding image design concepts CombinatorX generated that could fulfil the actual functions and meet the structure design requirements. Thus, the model can be a combination of metaphorical or literal elements of the additive 'love' without affecting the functionality of the chair. The designed model can be subjected to an iterative loop of testers to obtain desired results. Upon completion of each modelling session, participants will be required to refer to the combined text and image concepts generated by CombinatorX in order to self-evaluate the 3D model and then make modifications based on their expectation. After the modelling was completed, a semi-structured interview was conducted with the two participants to investigate whether the combinational concepts generated by the proposed method were useful for the actual design. The interview questions are included in the appendix B.

5.4. Results

5.4.1. With CombinatorX, the output was more than with the others

When using CombinatorX, participants generated an average of six corresponding images. However, when employing Internet searching, sketching and using Combinator, the participants both selected only four images that were deemed satisfactory. Design images generated using the three tools are illustrated in Figure 9. Given that the design topics

(a) A chair symbolizing love generated by CombinatorX

(b) An amalgamated image produced by Combinator

(c) A hand-drawn representation of love shaped chair

Figure 9. Design images generated using the three tools, with taglines of (a) A chair symbolising love, shaped like a flower, generated by CombinatorX, further enhanced with soft scene elements (b) An amalgamated image of a chair and love produced by Combinator (c) A hand-drawn representation by a user, featuring a chair with a backrest resembling a heart.

involved abstract terms, two participants expressed that relying solely on search engines made it challenging to conceive high-quality creative ideas, since the search outcomes were often overly specific and could constrain creative thinking. Additionally, due to time constraints, realising some intricate ideas became difficult. Although Combinator's operational process was straightforward, the space for creative exploration was somewhat limited. In contrast, CombinatorX offered a diverse array of embodiments for selection, and its ability to efficiently transform text into images significantly reduced the time required for illustration, ensuring higher quality results.

5.4.2. Combinatorx performance

Results of the NASA-TLX questionnaire are shown in Table 1. Participants observed that the mental and physical demands associated with CombinatorX were significantly lower, averaging scores of 6.5 and 2.5, respectively. In contrast, non-AL tools had average scores of 13.5 and 11.5. With non-AL tools, users were obligated to manually search for information related to keywords and ponder exhaustively on how to amalgamate these keywords. Conversely, within CombinatorX, the system generated a plethora of high-quality information, directly presenting users with embodiments, combination strategies and synthesised images. Consistent with these findings, the perceived effort and temporal demands for CombinatorX were also significantly lower than those for non-AI tools, scoring 9.5 and 5 respectively. Although the mental demand and physical demand of CombinatorX were similar to those of Combinator, its performance score (15.5) was considerably superior to Combinator's score of 6.

5.4.3. Users' feedback in CombinatorX

During stage 1, the embodiments proposed by the system are often perceived as valuable by users. Taking 'love' as an illustrative example, one user articulated: 'My preliminary design notions were predominantly shape-oriented. However, the CombinatorX unveiled

Table 1. NASA-TLX questionnaire results comparing CombinatorX, Combinator and Non-AI Tools.

	CombinatorX	Combinator	Internet Searching and Sketching	Inter-rater
Mental Demand	6.5	4.5	13.5	0.90
Physical Demand	2.5	2	11.5	0.98
Temporal Demand	5	2.5	15.5	0.98
Performance	15.5	6	13	0.95
Effort	9.5	5.5	13	0.98
Frustration	4	5.5	8.5	0.95

In the table, a downward arrow indicates that a lower value for the given metric is preferable, while an upward arrow suggests that a higher value is more desirable.

Figure 10. The 3D model of the combination 'chair + love' in modelling task by the participants using CombinatorX.

multiple imaginative directions for me: flames symbolising fervent and enduring passion, flowers representing pure affection, and rings signifying eternal and unending love'.

Transitioning to stage 2, the system capably visualises the design ideas from the prior phase. User feedback affirmed that this feature markedly enhanced their productivity, with the generated image quality garnering commendations. One user noted, 'Though there might be slight discrepancies in certain details, the images generated by the system align closely with the design theme, facilitating the fulfilment of the design task'. Regarding the scene generator, there is a consensus among users that it amplifies the overall visual appeal of the product imagery. For instance, a chair epitomising 'love' with a floral design, when complemented with the 'cushion' scene element, exudes an augmented sense of softness and warmth.

In the modelling task, the interviewees provided insights into the usefulness and necessity of conceptual reference materials in the design modelling process. The design models are shown in Figure 10. Firstly, participants generally agreed that although the combinational ideas provided could not be directly used in design implementation, text descriptions and visuals were still integral to the design modelling process. The generated design concept provides the basis for the interpretation of the loving form. An interviewee mentioned, 'Without the guidance of those textual and visual concepts, modelling would have been more difficult'. Second, additional detailing, such as the soft touch of the chair, was considered a key element in understanding and communicating the design concept 'love', allowing the design to demonstrate utility as well as convey a sense of warmth and intimacy.

6. Discussion

6.1. Discussion of approach evaluation

Stage 1 is the ideation stage where the LLM determines the combination of the additive concepts. In this stage, AI reasoning obtains higher novelty scores. We suggest that this difference is because the LLM, with its vast knowledge and information, is better at retrieving unfamiliar concepts than humans in the embodiments searching (White et al. 2023). The novelty of Stage 2 with the AI assistance was improved but the difference was not significant. One of the reasons may be that stage 2 does not involve exploring new combinational embodiments, but rather interpreting combinational concepts through visualisation (Koronis, Casakin, and Silva 2023). Such incremental improvement may hinder the novelty of ideas. In addition, considering that the training data of foundation models is based on real world knowledge (Göring et al. 2023), generating content that does not exist in world knowledge is challenging.

Our results show that the Strategy + AI group received the highest score, compared to both the Strategy Only group and the AI Only group. This indicates that combinational strategy and AI are synergistic in facilitating the combinational creativity. Particularly in terms of quality, quantity and variety, Strategy + AI approach is significantly better compared to the other two approaches. Upon closer examination, however, we found that the novelty scores for Strategy + AI were about the same as those for AI Only, both at the stage of concept verbalisation and externalisation. One possible explanation for this result is that the retrieval of relevant conceptual knowledge by the large foundation model is insensitive to prompts with combinational strategies (Guan et al. 2023) and tends to generate concept knowledge that is familiar and widely recognised by humans (Suresh et al. 2023). Novelty represents an unusualness compared to most existing concepts and its position is not flat in the conceptual space. In contrast, the training data for the large foundation model is derived from human knowledge and the model reasoning tends to be in regions that are flat in the conceptual space. This is because large-scale generative models are trained on vast datasets from the internet. Even when deploying prompt engineering to guide the models towards more accurate tasks, the vast knowledge space and numerous non-aligned data may lead models that may misinterpret or generate irrelevant content (Wang et al. 2023a).

For the comparison of Strategy Only and AI Only, the group of Strategy Only was an approach without AI involvement and received the lowest score in novelty, quality and quantity in overall. Notably, the variety score for Strategy Only was 62.07% higher than for AI Only. One possible reason is that in the context of combinatorial creativity, the ability of AI to synthesise two concepts can be beyond that of humans, and AI can effectively serve as an augmentation technique for combining concepts to assist synthesis. However, it seems that the AI still tends to generate those solutions that it considers most reasonable in terms of exploring multiple possibilities for combinations (Suresh et al. 2023). Although the combinational strategy does not have a direct impact on synthesis effects, it can be a way to help explore diverse combinations and expand thinking during the divergence and convergence phases (Childs et al. 2022).

6.2. Discussion of case study

The case study focuses on investigating the performance of the CombinatorX based on our approach in a practical combinational creativity design task, comparing with current mainstream combinational creativity tool, and examining the advantages and disadvantages of CombinatorX compared to autonomous human design. In addition, the case study also determines the usefulness of the generated ideas by the approach in facilitating actual combinational design practice.

The evaluation results show that CombinatorX received the highest score for combinational performance. It indicates that our approach is indeed effective in enhancing combinational creativity compared to human design and existing combinational creativity tools. Although CombinatorX is not as easy to use or as efficient as programmatic combinational tool, it significantly reduces labour in terms of mental, physical and temporal demands (51.85%, 78.26% and 67.74%, respectively) compared to human-autonomous design. In addition, CombinatorX resulted in the lowest rate of frustration (27.27% compared to Combinator and 52.94% compared to Internet searching and sketching). One possible explanation is that the low frustration stems mainly from the feedbacks that they have quick access to high-quality combination ideas. This suggests that the combinations generated based on combinational strategy and foundation model are of high quality and meet or exceed user's expectations of the combined results. The approach may provide users with positive feedback on generation and facilitate the refinement of combinational ideation.

Traditional engineering design modelling usually requires a reference that can materialise the concept (Jackson and Keefe 2016). The modelling of conceptual blending is challenging (Eppe et al. 2018). By reviewing the user feedback in the case study, we found that the image-text concepts generated by CombinatorX can effectively assist designers in combinational design modelling. Additionally, we observed that designers do not tend to design a model exactly like the reference image, but infer some additive elements of the image for inspiration to be more practical in iterations. These conceptual stimuli seem useful for modelling design iterations and loops. We believe that the appearance would be more beneficial for subsequent design such as advertisements and promotions. We suggest that the combinational concepts inferred and externalised by the proposed approach based on AI foundation model can be used as effective design references to accelerate practical engineering design tasks such as product design and industrial design.

6.3. Integrating the foundation model knowledge with human knowledge

While integrating AI and human knowledge to support design is quite novel, the design fixation caused by referencing generated ideas is worth considering (Jansson and Smith 1991). In our experiments, designers who referenced images generated by CombinatorX were not limited to designing the same model as the image. Instead, they preferred to iterate and improve their designs during the referencing process. The results of our study are consistent with those reported by the previous study (Lee and Chiu 2023). One possible explanation is that CombinatorX provides designers with a progressive usage of divergent reasoning and convergent iteration for concept generation, where the generated ideas do not lead to a design fixation.

While the knowledge space of foundation models far exceeds that of human individuals, not all useful knowledge can be recognised by such models in a specific creative task. In our experiments, Strategy Only and AI Only approaches performed difference between the abstract case and the concrete case. In the tasks involving human participants, the abstract combination 'chair + love' yielded superior results compared to the concrete combination of 'cup + bell'. However, the machine-involved combination shows the opposite result the machine scored higher on the concrete combination than the abstract one. This suggests that the machine has difficulty translating abstract concepts into concrete associations related to the Additive for combination.

A plausible explanation might be that AI foundation models tend to focus on some undisputed common-sense knowledge (Koralus and Wang-Maścianica 2023), and the Internet lacks such uncommon, combined text or images, which leads to a lack of relevant knowledge for the model. The model invariably offers what it perceives to be the optimal solution, even if such solutions are incorrect. One future direction is to integrate a novel search pathway for AI models to infer concrete associations from abstract concepts, such as integrating metaphorical reasoning or analogical reasoning (Helman 2013) to AI models. We believe that the expansive knowledge of large models can help humans gain a larger space for conceptual exploration, while the knowledge of human individuals can effectively guide the models to constrain the irrationality that may be caused by themselves.

6.4. Limitations and future work

Our current work has some limitations. Although the 'embodiments' we have defined cover the six primary features expected in the Additive, it does not account for all potential features. Future research can investigate categorising features and explore if other features also have an influence on the generated results. ChatGPT generates content we expect based on the keywords and prompts we provide. Considering the impact of prompt engineering on LLM and T2I, different keywords and prompts may affect the generated results, which may lead to text or image combination failure. Low-quality images cannot be used directly in subsequent design loops, and designers have insufficient control when modifying those low-quality images with prompts. Future research can consider identifying more structured and guided prompting means to support combinational creativity strategies. In addition, the proposed approach is to give the base rather than inviting the user to define their own. However, how to give and allow users to find additives is a challenge (Chen et al. 2019; Han et al. 2018). A potential direction of future work is exploring how to find combinational components by a data-driven approach.

The current version of CombinatorX focuses on product design for everyday life. However, for some product designs with complex systems, such as designing embedded devices with electronic motherboards, CombinatorX is challenging, especially in the part of externalisation stage. It is difficult for T2I to transform textual content into extremely detailed visual elements. Furthermore, CombinatorX is targeted at novice designers only. It is unclear whether professional designers can benefit from CombinatorX and to what extent it may enhance their creativity. In the design loop, we only considered iterations of design modelling outside the AI generation process. There is potential for integrating AI foundation model with design loop in both stage of verbalisation and externalisation. In the future, we aim to consider the development of a combinational creativity tool for the professional designer community. We believe that their conceptual discovery patterns, use cases and interaction modes will be very different from CombinatorX.

7. Conclusion

This study presents an AI foundation model enhanced approach for supporting combinational product design. The approach associates distinct concepts, suggests verbal combinations, and externalises visual images. It integrates a large language model and a text-to-image model. The language model supports divergent thinking and verbalisation, while the text-to-image model aids in the externalisation of visual concepts. Our experiment shows that the system, integrating combinational creativity strategies and AI, demonstrates robust and high-quality generative capabilities. The approach assists users in creating more and higher-quality ideas and reduces labour demand compared to traditional creative design approaches and automated combinational creativity tools. The effectiveness of the approach, the roles of combinational strategy and AI, and human-AI knowledge integration in the context of creativity support are discussed, which demonstrates the unique value of fusing creativity theory with AI.

Disclosure statement

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

Funding

This research is supported by National Key R&D Program of China (2022YFB3303304) and National Natural Science Foundation of China (62207023).

References

Bang, Y., S. Cahyawijaya, N. Lee, W. Dai, D. Su, B. Wilie, H. Lovenia, et al. 2023. “A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity.” In *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, 675–718.

Bar-Cohen, Y. 2006. “Biomimetics—Using Nature to Inspire Human Innovation.” *Bioinspiration & Biomimetics* 1 (1): P1–P12. https://doi.org/10.1088/1748-3182/1/1/P01.

Bender, E. M., T. Gebru, A. McMillan-Major, and S. Shmitchell. 2021. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. FAccT ‘21: 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event Canada*, 610–623. New York: ACM. https://doi.org/10.1145/3442188.3445922.

Bian, N., X. Han, L. Sun, H. Lin, Y. Lu, B. He, S. Jiang, and B. Dong. 2023. “ChatGPT is a Knowledgeable but Inexperienced Solver: An Investigation of Commonsense Problem in Large Language Models.” arXiv. Accessed December 29, 2023. http://arxiv.org/abs/2303.16421.

Boden, M. A. 2004. *The Creative Mind: Myths and Mechanisms*. Brighton, UK: Psychology Press.

Bonnardel, N., and E. Marmèche. 2004. “Evocation Processes by Novice and Expert Designers: Towards Stimulating Analogical Thinking.” *Creativity and Innovation Management* 13 (3): 176–186. https://doi.org/10.1111/j.0963-1690.2004.00307.x.

Brown, T., B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, et al. 2020. "Language Models are Few-Shot Learners." In *Advances in Neural Information Processing Systems*, 1877–1901. Vancouver, Canada: Curran Associates. Accessed December 18, 2023. https://papers.nips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html.

Chen, L., P. Wang, H. Dong, F. Shi, J. Han, Y. Guo, P. R. N. Childs, et al. 2019. "An Artificial Intelligence Based Data-Driven Approach for Design Ideation." *Journal of Visual Communication and Image Representation* 61:10–22. https://doi.org/10.1016/j.jvcir.2019.02.009.

Childs, P., J. Han, L. Chen, P. Jiang, P. Wang, D. Park, Y. Yin, et al. 2022. "The Creativity Diamond—A Framework to Aid Creativity." *Journal of Intelligence* 10 (4): 73. https://doi.org/10.3390/jintelligence10040073.

Chowdhery, A., S. Narang, J. Devlin, M. Bosma, G. Mishra, A. Roberts, P. Barham, et al. 2023. "Palm: Scaling Language Modeling with Pathways." *Journal of Machine Learning Research* 24 (240): 1–113.

Craft, A., B. Jeffrey, and M. Leibling. eds. 2001. *Creativity in Education*. King's Lynn, Norfolk: A&C Black.

Devlin, J., M. W. Chang, K. Lee, and K. Toutanova. 2019. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv. Accessed December 18, 2023. http://arxiv.org/abs/1810.04805.

Diao, S., P. Wang, Y. Lin, and T. Zhang. 2023. "Active Prompting with Chain-of-Thought for Large Language Models." arXiv. Accessed May 4, 2023. http://arxiv.org/abs/2302.12246.

Eppe, M., E. Maclean, R. Confalonieri, O. Kutz, M. Schorlemmer, E. Plaza, and K.-U. Kühnberger. 2018. "A Computational Framework for Conceptual Blending." *Artificial Intelligence* 256:105–129. https://doi.org/10.1016/j.artint.2017.11.005.

Eugeni, Q. 2020. *Love Chair*, *Reddot Design*. https://www.red-dot.org/project/love-44889.

Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. 2014. "Generative Adversarial Nets." In *Advances in Neural Information Processing Systems*, 2672–2680. Montreal: Curran Associates. Accessed August 15, 2023. https://proceedings.neurips.cc/paper_files/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html.

Göring, S., R. R. Ramachandra Rao, R. Merten, and A Raake. 2023. "Analysis of Appeal for Realistic AI-Generated Photos." *IEEE Access* 11:38999–39012. https://doi.org/10.1109/ACCESS.2023.3267968.

Guan, J., J. Dodge, D. Wadden, M. Huang, and H. Peng. 2023. "Language Models Hallucinate, but May Excel at Fact Verification." arXiv. Accessed December 29, 2023. http://arxiv.org/abs/2310.14564.

Han, J., M. Hua, D. Park, P. Wang, and P. R. N. Childs. 2020. "Computational Conceptual Distances in Combinational Creativity." In *Proceedings of the Design Society: DESIGN Conference*, 1, 177–186. https://doi.org/10.1017/dsd.2020.36.

Han, J., F. Shi, L. Chen, and P. R. N. Childs. 2018. "The Combinator – A Computer-Based Tool for Creative Idea Generation Based on a Simulation Approach." *Design Science* 4: e11. https://doi.org/10.1017/dsj.2018.7.

Helman, D. H. 2013. *Analogical Reasoning: Perspectives of Artificial Intelligence, Cognitive Science, and Philosophy*. Cleveland, USA: Springer Science & Business Media.

Ho, J., A. Jain, and P. Abbeel. 2020. "Denoising Diffusion Probabilistic Models." In *Advances in Neural Information Processing Systems*, 6840–6851. Curran Associates Inc. Red Hook, NY: Curran Associates. Accessed December 29, 2023. https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html.

Huang, D., and F. Guo. 2019. "Multiplicity of Periodic Bouncing Solutions for Generalized Impact Hamiltonian Systems." *Boundary Value Problems* 2019 (1): 57. https://doi.org/10.1186/s13661-019-1169-1.

Issa, L., F. Alghanim, and N. Obeid. 2019. *Computational Creativity: The Design of a Creative Computer Program*, 198. https://doi.org/10.1109/IACS.2019.8809107.

Jackson, B., and D. F. Keefe. 2016. "Lift-Off: Using Reference Imagery and Freehand Sketching to Create 3D Models in VR." *IEEE Transactions on Visualization and Computer Graphics* 22 (4): 1442–1451. https://doi.org/10.1109/TVCG.2016.2518099.

Jansson, D. G., and S. M. Smith. 1991. "Design Fixation." *Design Studies* 12 (1): 3–11. https://doi.org/10.1016/0142-694X(91)90003-F.

Kaufman, J. C., and R. J. Sternberg. 2010. *The Cambridge Handbook of Creativity*. San Bernardino: Cambridge University Press.

Keith, L. 2016. *Baby Bottle Drying Rack organicKidz Baby Bottle Tree*, *Reddot Design*. https://www.red-dot.org/project/organickidz-baby-bottle-tree-8716.

Ken Okuyama Design Co., Ltd. 2016. *Teaware ORIGAMI*, *Reddot Design*. https://www.red-dot.org/project/origami-33710.

Kingma, D. P., and M. Welling. 2019. "An Introduction to Variational Autoencoders." *Foundations and Trends® in Machine Learning* 12 (4): 307–392. https://doi.org/10.1561/2200000056.

Koizumi, S., and O. Yuichiro. 2021. *Kanpai Bell Pair*, *IF Design*. https://www.stylepark.com/de/news/if-award-gastlichkeit-tischkultur-kueche.

Koralus, P., and V. Wang-Maścianica. 2023. "Humans in Humans Out: On GPT Converging Toward Common Sense in both Success and Failure." arXiv. Accessed December 29, 2023. http://arxiv.org/abs/2303.17276.

Koronis, G., H. Casakin, and A. Silva. 2023. "An Experimental Comparison of Analogy Representation Effects on Creative Outcomes." *The Journal of Creative Behavior* 57 (4): 711–729. https://doi.org/10.1002/jocb.611.

Lee, Y.-H., and C.-Y. Chiu. 2023. "The Impact of AI Text-to-Image Generator on Product Styling Design." In *Human Interface and the Management of Information*, edited by H. Mori and Y. Asahi, 502–515. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-35132-7_38.

Liu, Yang, E. Jun, Q. Li, and J. Heer. 2019a. "Latent Space Cartography: Visual Analysis of Vector Space Embeddings." *Computer Graphics Forum* 38 (3): 67–78. https://doi.org/10.1111/cgf.13672.

Liu, Yinhan, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, et al. 2019b. "RoBERTa: A Robustly Optimized BERT Pretraining Approach." arXiv. Accessed December 18, 2023. http://arxiv.org/abs/1907.11692.

Liu, V., H. Qiao, and L. Chilton. 2022. "Opal: Multimodal Image Generation for News Illustration." In *UIST '22: The 35th Annual ACM Symposium on User Interface Software and Technology*, 1–17. Bend, OR: ACM. https://doi.org/10.1145/3526113.3545621.

Lloyd-Cox, J., A. P. Christensen, P. J. Silvia, and R. E. Beaty. 2021. "Seeing Outside the Box: Salient Associations Disrupt Visual Idea Generation." *Psychology of Aesthetics, Creativity, and the Arts* 15 (4): 575–583. https://doi.org/10.1037/aca0000371.

Marcus, G. 2020. "The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence." arXiv. Accessed April 3, 2024. http://arxiv.org/abs/2002.06177.

Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. 2013. "Distributed Representations of Words and Phrases and Their Compositionality." In *Advances in Neural Information Processing Systems*, 3111–3119. Mountain View: Curran Associates. Accessed December 18, 2023. https://papers.nips.cc/paper_files/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html.

OpenAI. 2023. "OpenAI Terms of Use." Accessed April 3, 2024. https://openai.com/policies/terms-of-use (Accessed: 3 April 2024).

Pavlichenko, N., and D. Ustalov. 2023. "Best Prompts for Text-to-Image Models and How to Find Them." In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2067–2071. https://doi.org/10.1145/3539618.3592000.

Radford, A., J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever. 2019. "Language Models are Unsupervised Multitask Learners." Accessed December 18, 2023. https://www.semanticscholar.org/paper/Language-Models-are-Unsupervised-Multitask-Learners-Radford-Wu/9405cc0d6169988371b2755e573cc28650d14dfe.

Raffel, C., N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, et al. 2020. "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer." *Journal of Machine Learning Research* 21 (140): 1–67.

Ramesh, A., P. Dhariwal, A. Nichol, C. Chu, and M. Chen. 2022. "Hierarchical Text-Conditional Image Generation with CLIP Latents." arXiv. https://doi.org/10.48550/arXiv.2204.06125.

Ramesh, A., M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever. 2021. "Zero-Shot Text-to-Image Generation." In *Proceedings of the 38th International Conference on Machine Learning*, PMLR, 8821–8831. Accessed August 15, 2023. https://proceedings.mlr.press/v139/ramesh21a.html.

Rombach, R., A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. 2022. "High-Resolution Image Synthesis with Latent Diffusion Models." In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10674–10685. New Orleans, LA: IEEE. https://doi.org/10.1109/CVPR52688.2022.01042.

Saharia, C., W. Chan, S. Saxena, L. Li, J. Whang, E. L. Denton, K. Ghasemipour, et al. 2022. "Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding." *Advances in Neural Information Processing Systems* 35:36479–36494.

Shah, J. J., S. M. Smith, and N. Vargas-Hernandez. 2003. "Metrics for Measuring Ideation Effectiveness." *Design Studies* 24 (2): 111–134. https://doi.org/10.1016/S0142-694X(02)00034-0.

Shneiderman, B. 2007. "Creativity Support Tools: Accelerating Discovery and Innovation." *Communications of the ACM* 50 (12): 20–32. https://doi.org/10.1145/1323688.1323689.

Suresh, S., K. Mukherjee, X. Yu, W.-C. Huang, L. Padua, and T. Rogers. 2023. "Conceptual Structure Coheres in Human Cognition but not in Large Language Models." In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. EMNLP 2023*, edited by H. Bouamor, J. Pino, and K. Bali, 722–738. Singapore: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.47.

Tan, L., and M. Luhrs. 2024. "Using Generative AI Midjourney to Enhance Divergent and Convergent Thinking in an Architect's Creative Design Process." *The Design Journal*, 1–23. https://doi.org/10.1080/14606925.2024.2353479.

van den Oord, A., S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, et al. (2016) "WaveNet: A Generative Model for Raw Audio." arXiv. Accessed September 6, 2023. http://arxiv.org/abs/1609.03499.

Veale, T. 2019. "From Conceptual Mash-Ups to Badass Blends: A Robust Computational Model of Conceptual Blending." In *Computational Creativity*, edited by T. Veale and F. A. Cardoso, 71–89. Cham: Springer International Publishing (Computational Synthesis and Creative Systems). https://doi.org/10.1007/978-3-319-43610-4_4.

Wallace, B., C. P. Martin, J. Tørresen, and K. Nymoen. 2021. "Exploring the Effect of Sampling Strategy on Movement Generation with Generative Neural Networks." In *Artificial Intelligence in Music, Sound, Art and Design*, edited by J. Romero, T. Martins, and N. Rodríguez-Fernández, 344–359. Cham: Springer International Publishing. Lecture Notes in Computer Science. https://doi.org/10.1007/978-3-030-72914-1_23.

Wang, S., S. Petridis, T. Kwon, X. Ma, and L. B. Chilton. 2023b. "PopBlends: Strategies for Conceptual Blending with Large Language Models." In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–19. Hamburg: ACM. https://doi.org/10.1145/3544548.3580948.

Wang, Y., W. Zhong, L. Li, F. Mi, X. Zeng, W. Huang, L. Shang, et al. 2023a. "Aligning Large Language Models with Human: A Survey." arXiv. Accessed December 29, 2023. http://arxiv.org/abs/2307.12966.

Wang, B., Y. Zhu, L. Chen, J. Liu, L. Sun, and P. Childs. 2023c. "A Study of the Evaluation Metrics for Generative Images Containing Combinational Creativity." *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 37: e11. https://doi.org/10.1017/S0890060423000069.

Wei-Chieh, L., and L. Hung-Hui. 2023. *SADDLE Chair/Lounge Chair*, *IF Design*. https://ifdesign.com/en/winner-ranking/project/saddle/562489.

Weisz, J. D., M. Muller, S. Houde, J. Richards, S. I. Ross, F. Martinez, M. Agarwal, and K. Talamadupula. 2021. "Perfection Not Required? Human-AI Partnerships in Code Translation." In *26th International Conference on Intelligent User Interfaces*, 402–412. https://doi.org/10.1145/3397481.3450656.

White, J., Q. Fu, S. Hays, M. Sandborn, C. Olea, H. Gilbert, A. Elnashar, et al. 2023. "A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT." arXiv. Accessed April 21, 2023. http://arxiv.org/abs/2302.11382.

Wu, X., and L. Li. 2024. "An Application of Generative AI for Knitted Textile Design in Fashion." *The Design Journal* 27 (2): 270–290. https://doi.org/10.1080/14606925.2024.2303236.

Yang, J., H. Jin, R. Tang, X. Han, Q. Feng, H. Jiang, S. Zhong, et al. 2023. "Harnessing the Power of llms in Practice: A Survey on Chatgpt and Beyond." *ACM Transactions on Knowledge Discovery from Data* [Preprint].

Zhang, M., and J. Li. 2021. "A Commentary of GPT-3 in MIT Technology Review 2021." *Fundamental Research* 1 (6): 831–833. https://doi.org/10.1016/j.fmre.2021.11.011.
Zhang, S., S. Roller, N. Goyal, M. Artetxe, M. Chen, S. Chen, C. Dewan, et al. 2022. "OPT: Open Pre-trained Transformer Language Models." arXiv. https://doi.org/10.48550/arXiv.2205.01068.
Zuo, H., Q. Jing, T. Song, L. Sun, P. Childs, and L. Chen. 2022. "WikiLink: An Encyclopedia-Based Semantic Network for Design Creativity." *Journal of Intelligence* 10 (4): 103. https://doi.org/10.3390/jintelligence10040103.

Appendices

Appendix A: interview survey: exploring the experiences in AI generative combinational design

This appendix contains the interview questions designed to explore designers' experiences and feelings during the product testing phase.

Thank you for participating in our interview. Your insights are invaluable to us as we seek to understand the experiences of designers during and after the product design testing phase. The purpose of this interview is to gather your thoughts, feelings, and observations on the AI generative design testing process, its challenges, and its outcomes. Your feedback will help us improve and innovate in our design approaches and methodologies.

Q1) 'Did you find it challenging to search for embodiments? If so, why?'

Q2) 'Did you encounter any difficulty in combining the embodiments with the Base?'

Q3) 'Did the resulting combination meet your expectations? Does the textual description comprehensively capture the features of both the Base and the Additive?'

Q4) 'Did you find it challenging to execute the visualisation of combined images? Why or why not?'

Q5) 'How would you rate the quality of the visualised image combination? Does it align with your ideal representation?'

Q6) 'Did the image consistently incorporate features from both the Base and the Additive?'

Appendix B: interview survey: user feedback on CombinatorX

Creative design modelling task self-Assessment interview 1

Q1) 'How similar do you think the post-modelling design model is to the provided image, in terms of shape and function?'

The similarity is moderate. The provided image integrates the shape of a heart somewhat rigidly, and its functionality lacks comfort. The design model has improved in both shape and function, with a smoother heart shape that aligns with human curves, offering greater practicality.

Q2) 'How helpful was generating this image for your modelling? Would the absence of this image have increased the difficulty of your modelling? (How helpful was the text-to-image process?)'

It was somewhat helpful, particularly in understanding the concept of the heart shape. Without the image, there would be some difficulty in modelling.

Q3) 'The image includes additional treatment to the chair, adding a soft and warm touch. Does this delicate softness help you better understand the design concept of the "heart chair"?'

Yes, it helps. Thus, in the design, we continued this smooth treatment, conveying a sense of friendliness and warmth without sharp edges.

Q4) 'Do you think your modelling is a good design for the "heart chair" task (is this modelling suitable for production)?'

It is a fairly good design, incorporating elements of the 'heart chair' into the current chair design, making it practical and creative.

Creative design modelling task self-Assessment interview 2

Q1) 'How similar do you think the post-modelling design model is to the provided image, in terms of shape and function?'

The similarity in shape is decent, and in terms of function, it has been shaped into a chair, so there is a certain degree of similarity; overall, it is highly similar.

Q2) 'How helpful was generating this image for your modelling? Would the absence of this image have increased the difficulty of your modelling?'

It provided a physical reference which is crucial, especially in traditional mechanical modelling which often requires a reference. Having such an image helps with the direction of modelling and doesn't significantly increase the difficulty.

Q3) 'The image includes additional treatment to the chair, adding a soft and warm touch. Does this delicate softness help you better understand the design concept of the "heart chair"?'

I have a certain understanding; it feels like a chair with a strong cyberpunk vibe.

Q4) 'Do you think your modelling is a good design for the "heart chair" task (is this modelling suitable for production)?'

I feel my modelling is too industrial, as I used SolidWorks for mechanical modelling, which might not align perfectly with the softness of the 'heart chair'.