Highlights

- A novel computational conceptual design approach using LLMs.
- Supports designers in analogy-based structured retrieval for ideation.
- Facilitates reasoning and synthesis through morphological and semantic mappings.
- Introduces ViMimic, an interactive tool for creative conceptual design.
- Evaluated for creativity enhancement, efficiency, and cognitive load reduction.

From Analogy to Innovation: A Creative Conceptual Design Approach Leveraging Large Language Models

Boheng Wang[a], Xiaoyang Zhao[a], Haoyu Zuo[a], Yaxuan Song[b], Ji Han[c], Peter Childs[a], Liuqing Chen[b,*]

[a]*Dyson School of Design Engineering, Imperial College London, Exhibition Rd, South Kensington, SW7 2AZ, United Kingdom*
[b]*College of Computer Science and Technology, Zhejiang University, Hangzhou, 310030, China*
[c]*Department of Innovation, Technology and Entrepreneurship, University of Exeter, Exeter, EX4 4RN, United Kingdom*

Abstract

Integrating creative concepts into Product Design and Manufacturing Systems (PDMS) is important for product innovation. However, current PDMS lack cognitive capabilities, particularly in reasoning and synthesis, which are essential for conceptual design. As a result, designers face challenges in retrieving relevant analogies, establishing meaningful mappings, and integrating knowledge into new design concepts. This paper proposes a computational conceptual product design approach that integrates Large Language Models' (LLMs) knowledge representation with an analogy-based structured retrieval mechanism, supporting designers to explore and recombine design patterns and functionalities in an intuitive manner. Benefiting from the zero-shot learning and prompting capabilities of LLMs, given a source domain, this approach allows reasoning target domain based on abstract correspondences in both morphological and semantic associations. A template for the combinational regulation of the reuse of analogical knowledge has also been formulated. By decomposing analogical knowledge into ontological distinction, inspirational feature recognition, and associative mapping explanation, a creative conceptual design stimulation path is formed. An interactive tool named ViMimic based on this approach has been developed through a case

*Corresponding author
Email address: `chenlq@zju.edu.cn` (Liuqing Chen)

study with 18 participants. Evaluation results demonstrate that the approach improves creative performance, increasing the novelty and functionality of conceptual designs by 51% and 22% respectively according to expert evaluations. It also boosts the efficiency and diversity of analogy mapping by 30% based on objective measures, while enhancing creative experiences and reducing cognitive load as measured in the participants' self-assessment.

Keywords: Large Language Models, Conceptual Product Design, Analogy, Creativity, Innovation, Artificial Intelligence

---

1. Introduction

Integrating creative concepts into Product Design and Manufacturing Systems (PDMS) is important for product innovation. PDMS provide computational support for product development, enabling designers to manage structured processes such as modeling, simulation, and optimization. These systems play a key role in ensuring efficiency, accuracy, and consistency in design and manufacturing, helping industries improve workflows and product quality. Over the years, PDMS have evolved to incorporate advanced computational techniques, facilitating data-driven decision-making and automation in design processes (Baby and Nellippallil, 2024; Peruzzini and Pellicciari, 2017). However, while PDMS are effective in optimizing well-defined design tasks, it offers limited support for conceptual design, which is an essential phase where innovative ideas are generated before detailed development begins (Childs et al., 2022).

Conceptual design requires designers to explore diverse possibilities, identify novel configurations, and synthesize insights from different domains. This process involves exploring diverse sources of inspiration, one of which is analogy-based design (Goel, 1997), where knowledge from a source domain is mapped onto a target domain to inspire new solutions. By identifying structural or functional similarities between different concepts, designers can generate creative solutions (Gick and Holyoak, 1980). However, existing PDMS primarily focus on structured workflows and fixed computational processes, lacking mechanisms to facilitate analogical retrieval, mapping, and synthesis. This gap forces designers to manually search for relevant analogies, assess their applicability, and integrate them into new design concepts. Such a manual approach is time-intensive and dependent on individual expertise, reducing the scalability and efficiency of analogy-driven ideation within

PDMS.

Recent advancements in Large Language Models (LLMs) offer a potential solution for enhancing PDMS by enabling automated analogy retrieval and reasoning. LLMs contain broad cross-domain information and demonstrate the ability to identify relationships between concepts without direct training (Nooralahzadeh et al., 2020). However, their outputs often lack interpretability and structured mapping, making them difficult to integrate into systematic design workflows. Additionally, LLMs struggle with generating analogical insights that go beyond existing knowledge representations, restricting their ability to support open-ended innovation.

This paper proposes a computational conceptual product design approach that integrates LLMs with a structured analogy retrieval mechanism to enhance conceptual design in PDMS. The approach is composed of five key components: Problem Finding, Primitives Scheme Block (PSB), Analogy Mapping, Scoring Block, and Combination. The process begins with Problem Finding and PSB. Problem Finding receives an input, which is a keyword of a design task. PSB is a block where the design task is structured into a solution space and a decomposed problem space. The PSB extracts archetypal solutions, which are stable, well-defined concepts that serve as source domain references for analogy reasoning. These solutions, along with contextual information such as design scenarios and user needs, provide the foundation for analogy retrieval. The Analogy Mapping component utilizes LLMs to identify potential cross-domain analogies based on semantic and morphological associations. Since direct analogy retrieval can be inconsistent, a Scoring Block evaluates the generated analogies by measuring their conceptual relevance and structural similarity to ensure meaningful mappings. Finally, the Combination Block synthesizes the retrieved analogical knowledge into new design concepts, integrating insights from both the source and target domains. This structured process enables systematic exploration and reuse of analogical knowledge, supporting designers in generating novel and functional conceptual designs.

To validate this approach, an interactive tool named ViMimic was developed, which automates analogy retrieval, reasoning, and combination within PDMS. A case study demonstrates the effectiveness of the approach. The results show that the proposed approach enables diverse analogies, improves reasoning efficiency, reduces cognitive load, and enhances the innovativeness and creativity of conceptual design, by LLMs-based analogy mapping and domain knowledge reusing.

The remainder of this paper is organized as follows: Section 2 presents a brief review of analogy within the context of design, the current state of research on analogy-based design, and related works on leveraging large language models to drive design innovation, thereby identifying the research gaps. Section 3 provides a detailed explanation of the framework and technical specifics of the proposed approach. Section 4 demonstrates the effectiveness of this approach in supporting creative conceptual design through a case study. Sections 5 and Section 6 present the discussion and conclusion, respectively.

2. Related work

2.1. Analogy

Analogy is a fundamental cognitive mechanism that allows people to understand novel situations, solve problems, and generate creative insights by drawing parallels between seemingly unrelated domains (Gentner and Holyoak, 1997; Forbus et al., 1998). It enables individuals to map the relational structure of one situation onto another, facilitating abstraction, generalization, and flexible thinking (Gentner and Markman, 1997). The outcome of analogy often results in new cognitive connections or innovation, which can serve as a foundation for further creative thinking (Holyoak and Thagard, 1996). To achieve this, analogical combination is effective (Han et al., 2019a). When combining analogies, users often discover unexpected connections and points of convergence between the two. For example, people use the tree's branching structure to design efficient urban transportation networks, combining ideas between natural growth patterns and human infrastructure. This approach may facilitate novel associations that can yield unexpected insights, fostering designs that are both original and adaptable (Boden, 2009).

Given the importance of analogy in human cognition and innovation, researchers have sought to model analogical reasoning through algorithmic approaches to promote the development of computational intelligent systems. For example, generating inferences and learning about new domains by utilizing previous knowledge and experience (Daugherty and Mentzer, 2008), that is, applying knowledge from a source domain to a target domain (Blanchette and Dunbar, 2000). However, mapping knowledge between different fields is challenging. This involves the relationships between domains and the transfers of cross-domain knowledge. To address this issue, various theoretical frameworks and models have been proposed to define analogical distance. Far-distance analogy has been investigated to encourage designers to extend their focus from familiar domains to new ones, thereby discovering commonalities between seemingly unrelated entities (Han et al., 2019b). Some research argues that the most creative solutions may come from stimuli that are closer to the problem, because these stimuli are more relevant and perceptible in relation to the problem (Fu et al., 2013).

Despite the advantages of analogy, modeling analogical reasoning presents challenges such as limited flexibility in mapping structures, lack of cross-domain knowledge in using external databases, and difficulty handling analogical representations in data-driven workflows. Additionally, the effectiveness of analogy retrieval depends on cognitive flexibility, prior knowledge, and problem framing, necessitating strategies to optimize analogy selection and application.

2.2. Early stages of product design and manufacturing

Designers are frequently required to generate novel ideas, retrieve cross-domain knowledge, and synthesize information from disparate sources (Chandrasegaran et al., 2013). However, attention to Product Design and Manufacturing Systems for supporting such advanced cognitive functions has had limited attention to date. Traditional PDMS have focused on the efficient execution of engineering and manufacturing tasks (Koren et al., 2018). For example, existing systems primarily rely on parametric operators, which inherently limit their ability to align with and fully support the design process (Bettig et al., 2005). In addition, current PDMS tools are generally geared towards convergent tasks, focusing on refining existing designs rather than generating novel ideas (Boothroyd et al., 2010; Hoque et al., 2013). This focus limits the exploration of alternative design concepts. As a result, designers struggle to identify relevant precedents, uncover latent connections between design elements, and integrate creative insights into viable product concepts (Howard et al., 2008).

Although some data-driven decision-making algorithms (Cantamessa et al., 2020) and representation-based modelling (Gorti et al., 1998) have been proposed to support cognitive functions, they still lack mechanisms for understanding abstract concepts, reasoning about design requirements, and mapping conceptual ideas to potential solutions. Data-driven design algorithms learn from existing data – past designs, consumer behaviors, or simulation results – and thus tend to reproduce or tweak known solutions rather than invent fundamentally new ones (Johnson et al., 2023). In other words, an algorithm can efficiently find a better version of what already exists, but is less likely to generate a bold concept outside the scope of its training data. Beyond creativity, traditional cognitive PDMS have shown limitations in open-ended problem-solving and adaptation. These systems excel when the design problem is well-defined (with clear objectives, constraints, and plenty of relevant data). However, real-world design challenges are often ambiguous, with shifting goals and interdependent variables. Designers often approach problems by drawing analogies (i.e., how else has a similar challenge been solved) or by making intuitive hypotheses to explore (Goel and Bhatta, 2004). Conventional optimization algorithms in PDMS lack this flexible reasoning.

To address these issues, there is a need for PDMS to incorporate cognitive augmentation features that can support and scaffold early-stage design thinking. This includes natural language understanding, creative reasoning, and knowledge reuse.

2.3. Large language model for design innovation

Large Language Models are advanced AI systems based on deep learning that train artificial neural networks on vast and diverse datasets to understand and generate human-like language (Brown et al., 2020). Their knowledge space is constructed from extensive text corpora. This broad exposure allows LLMs to develop a generalized understanding of world knowledge (Yu et al., 2023b), encompassing topics such as history, science, engineering, and culture. Unlike traditional AI systems, which are often confined to specific datasets or narrow applications, LLMs leverage deep learning architectures—particularly Transformer (Vaswani et al., 2017) based models—to learn contextual relationships between words, concepts, and patterns. This enables them to retrieve, synthesize, and reason information with coherence and adaptability (Yang et al., 2022), making them capable of answering complex questions, summarizing content, and even offering creative insights. Some studies have explored LLM-driven innovation in the fields of engineering and design. Using LLM-generated content as creative concepts has become an emerging paradigm in the current engineering design community for creativity support (Zhu and Luo, 2022; Chen et al., 2024b; Ma et al., 2023). LLMs' exceptional performance, text-based input accessibility, and flexible APIs undoubtedly provide opportunities for researchers and providers in developing AI-driven tools to explore creative design. By referencing generated text, engineers and designers can gain inspiration, explore diversity, refine design ideas, or make decisions (Duan and Zhang, 2022; Kwon et al., 2023). A benefit is that AI significantly shortens the time for ideation, allowing creators to experiment and iterate quickly (Joosten et al., 2024; Vinichenko et al., 2020). Despite the promising prospects of generative AI in design innovation, some challenges have emerged. For example, current LLMs effectively capture common-sense knowledge but lack the ability to transform this knowledge into creative solutions (Bian et al., 2024). Additionally, the opacity of generative algorithms poses challenges for interpretability and accountability, making it difficult for designers to understand or justify AI-generated outcomes.

While the inference process of general models is relatively straightforward to implement, effectively controlling the generated output via prompts presents a significant challenge. Due to the lack of a prompting strategy and experience of users, models often return familiar-looking images or superficially plausible but flawed information (Dathathri et al., 2019). Users may end up with unsatisfactory results. Recent research has explored various prompting techniques to enhance the reasoning capabilities of LLMs. Chain-of-Thought prompting contributes by guiding the model through explicit reasoning steps, significantly enhancing performance on tasks that require logic, arithmetic, or multi-step reasoning (Wei et al., 2022). Retrieval-Augmented Generation (RAG) enriches prompting by incorporating external, non-parametric memory through document retrieval (Lewis et al., 2020). This makes RAG particularly effective for knowledge-intensive or time-sensitive tasks where pretrained models may lack up-to-date or domain-specific information. Few-shot prompting addresses this by including a handful of task examples, offering better calibration of model behavior and improved performance in scenarios requiring subtle contextual understanding (Ye and Durrett, 2022). Zero-shot prompting relies solely on natural language instructions without examples, providing a simple and scalable approach, especially useful in low-resource or time-sensitive settings (Brown et al., 2020). Although Zero-Shot Prompting may lack precision in complex engineering tasks, this imprecision can be a source of creative potential in conceptual design. A degree of vagueness can be advantageous in promoting creative exploration in the early phases of design (Mascitelli, 2000).

Based on the literature review, the current cognitive product design approaches have significant gaps in three key aspects: Firstly, the obtained analogy knowledge may not directly help users in creative ideation. Exist-

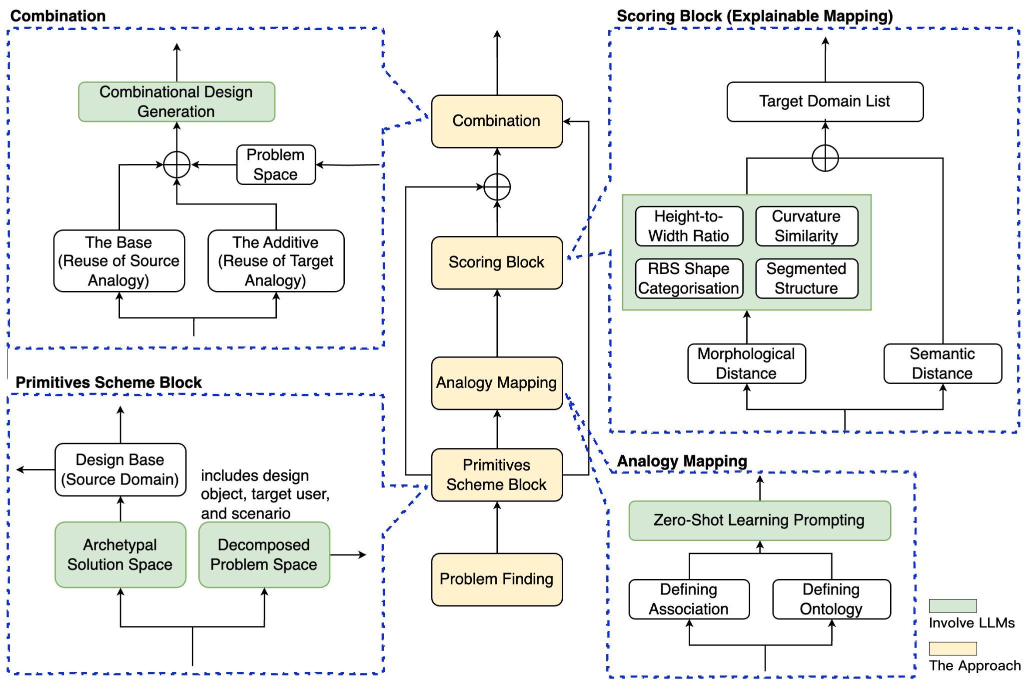

Fig. 1. The overall framework of the LLMs-based conceptual product design approach.

ing processes face difficulties in effectively transforming scattered analogical information into innovative design concepts, leading to a gap between knowledge and creativity. Secondly, current PDMS lack advanced cognitive augmentations to support conceptual reasoning. Finally, although LLMs have achieved scores above the human average in common-sense reasoning tasks, it is difficult for them to make significant breakthroughs in complex and inexperienced tasks (Bian et al., 2024), which hinders the efficient integration of cross-disciplinary knowledge and creative expression.

3. Methodology

This paper proposes a computational conceptual product design approach, combining large language models with a structured retrieval mechanism based on analogy. This promotes the creativity of conceptual products through reasoning and reusing analogy knowledge. As shown in Fig. 1, the overall framework of the approach consists of five blocks, including Problem Finding, Primitives Scheme Block (PSB), Analogy Mapping, Scoring Block, and

Combination. The statement of the design problem is taken as the initial input and finally generates a combinational creative conceptual design by LLMs. The green box in the diagram represents LLMs as artificial agents responsible for problem decomposition, analogy reasoning and evaluation, and generating creative product design in the stages of conceptual design generation. Throughout the process, the input is reasoned by the PSB to a solution space and a decomposed problem space, where these solutions are stable concepts existing in the human knowledge space and therefore may not be creative, and we refer to them as archetypal solutions. The analogy mapping block uses the archetypal solutions as source domain concepts for the LLM's automated analogy reasoning, where analogy association and analogy distance can be customized. Considering insufficient interpretability and controllability of the LLM in reasoning tasks, we designed a scoring block to evaluate and align the target domain analogy after analogy mapping, ensuring the effectiveness and stability of the black-box analogy based on LLM. Finally, the decomposed problem space information, target domain information, and source domain information from PSB are reused in the combination block to generate creative and novel conceptual designs. It is noteworthy that we do not involve external knowledge bases to participate in the entire approach. This is because the conceptualization of analogy and creativity often requires breaking the cognitive rules of inherent and static knowledge (Holyoak and Thagard, 1996). External knowledge bases may not fully support this flexible reasoning process and may limit the model's creativity and adaptability in the face of new problems. In addition, for mature pre-trained large language models, the scale of built-in knowledge has basically reached the level of world knowledge (Yu et al., 2023a). Rich contextual information and conceptual associations can be implicitly represented and captured by LLMs which are sensitive to input (Lee et al., 2024). Therefore, our focus is to control the entire process of LLM-driven analogy reasoning and knowledge reusing and fully utilize the knowledge space of pre-trained LLMs.

3.1. Primitives scheme block

The design goal originates from a problem requirement. Whether for a PDMS or a human designer or engineer, they both need external inputs or stimuli to begin analysis, understanding, and implementation of design. However, the description of design problems by users is often uncertain; they may be a clearly defined product function scheme or a vague requirement that includes target users, scenarios, and contextual information. This unstructured input cannot be well integrated into the computational system for analysis. Therefore, we use the LLM to parse the initial description into a basic scheme space and a problem space that decomposes additional scenarios, objectives, and user information, denoted as mapping M:

$$\mathcal{M} : D \rightarrow (S_a, P) \tag{1}$$

where D represents the initial description. The scheme space S_a is a set of conventional solutions to design problems, which have solidified main functional semantics, and these solutions have been tested historically extensively. We call this type of scheme as archetypal solution, denoted as $S_a = \{s_1, s_2, \ldots, s_n\}$. For example, an archetypal solution that provides comfortable support and sitting rest functions can usually be described as a sofa, chair, or stool. Although these solutions do not seem novel on the surface, they have stable and easily understandable functional solutions semantically. Thus, the S_a can be used as source domain objects for analogy and as basic concepts for creative design, participating in subsequent analogy reasoning and conceptual synthesis. The problem space P is represented by Eq. (2), consisting of three component sets decomposed by LLM, including design scenario $C = \{c_1, c_2, \ldots, c_k\}$, object $Obj = \{obj_1, obj_2, \ldots, obj_l\}$, and user information $U = \{u_1, u_2, \ldots, u_p\}$.

$$P = C \cup Obj \cup U \tag{2}$$

For example, the design scenario of a chair is product design, the target user is office workers, and the objective is to provide a massage and change sedentary habits. The parameters of the problem space are not necessary. According to user input, their values can be empty, i.e., for a more general design. Following this, the complete PSB can be expressed as a function f_{psb}:

$$(S_a, P) = f_{psb}(D) = (\{s_1, s_2, \ldots, s_n\}, C \cup Obj \cup U) \tag{3}$$

3.2. Generating and scoring analogies

3.2.1. Analogy mapping

In this block, two propositions need to be defined in advance to prepare for the automated analogy of the LLM, namely the associations and the ontology. According to structural mapping theory, the core of analogical reasoning refers to structural alignment, which means establishing this relationship between the source domain and the target domain, rather than simple attribute matching, which can be calculated by distance metrics. Therefore, analogy is defined as a mapping $M : S \rightarrow T$ from the source domain S to the target domain T, and its basic formula can be expressed as:

$$M : R_S(x, y, \ldots) \rightarrow R_T(x', y', \ldots),\ d(S, T) \tag{4}$$

where $S \in S_a$. $R_S(x, y, \ldots)$ is the relationship set in the source domain, and $R_T(x', y', \ldots)$ is the relationship set in the target domain. $d(S, T)$ is the analogical distance function between the source domain S and the target domain T, which are described in Section 3.2.2. Inspired by visual analogy (Yaner and Goel, 2006), two common analogy associations are considered, namely morphological and semantic associations. The reason for considering these is that the relationship in visual analogy is usually similar in overall form and structure, while significant differences in semantics. This relationship often leads to unexpected creative breakthroughs, allowing designers to consider problems outside the box (Gassmann and Zeschky, 2008).

Considering that this approach focuses on product design and manufacturing, we define "object" as the top-level concept of the ontology structure and differentiate it into artifacts and natural objects, based on their origin and manufacturing method. The reason for defining the ontology structure in this way is that the relationship between the two in design and manufacturing is unidirectional, that is, humans can refer existing natural objects to design artifacts, but cannot create natural objects. Therefore, new design ideas can be referenced and iterated from past engineering achievements as well as inspired by natural objects. This notion will be reflected in the combination block. By utilizing the Zero-Shot Learning ability of the LLMs and prompt engineering (Marvin et al., 2023), we can effectively retrieve potential analogy knowledge, denoted as Eq. (5),

$$(T, O) = f_{map}(p(S, R)),\ D_{train} = \emptyset \tag{5}$$

where f_{map} represents the LLMs as the analogy mapping functions, p represents the prompting wrapper, R refers to the analogical association, O represents the ontology category, and D_{train} represents trained samples. Considering that analogical association is essentially "jumping" relationship matching rather than step-by-step logical reasoning, examples may lead to inherent pattern matching. Therefore, in this study, we did not consider Few-Shot Prompting (Reynolds and McDonell, 2021), Chain-of-Thought (CoT) Prompting (Wei et al., 2022), and Retrieval-Augmented Generation (RAG)

(Lewis et al., 2020). These methods are more suitable for structured reasoning with clear steps or specific knowledge dependencies, such as TRIZ (Altshuller, 1984) and SCAMPER (Serrat and Serrat, 2017).

3.2.2. Scoring block

Although knowledge retrieval by prompting is an efficient means of utilizing LLMs, its black-box mechanism and inherent biases in the training data may amplify "hallucination" (Perković et al., 2024). The analogy results may appear reasonable on the surface but are actually incorrect, and it is impossible to know how they were inferred. This means that using a single analogy mapping block to perform analogical reasoning may not be reliable. To address this issue, a scoring block is proposed to evaluate the relationship distance between source domain objects and target domain objects before and after analogy. This allows sorting of analogies from high to low reliability, with analogies more highly ranked less likely to be hallucinations produced by the LLM. For language models, calculating semantic distance is very simple, and the distance can be represented as the direction of two word vectors in high-dimensional space, as shown in Eq. (6). $||\mathbf{v}_1||$ and $||\mathbf{v}_2||$ are the Euclidean norms of vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, respectively.

$$\cos(\mathbf{v}_1, \mathbf{v}_2) = \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{||\mathbf{v}_1||||\mathbf{v}_2||} \tag{6}$$

In Eq. (4), the distance function $d(S, T)$ can reflect the mapping relationship between objects from a computational perspective. In analogy systems, they can be near or far distances, and their values can be defined by different methods. For example, Eq. (7) represents the mismatch distance of relationships:

$$d(S, T) = 1 - \frac{|M_R|}{|R_S|} \tag{7}$$

where $|M_R|$ represents the number of successfully mapped relationships, and $|R_S|$ is the total number of relationships in the source domain. When all relationships match, $d(S, T) = 0$, indicating the most similarity; when no relationships match, $d(S, T) = 1$, indicating the most dissimilarity. By combining the cosine distance and normalization, it can be defined as:

$$d_s(S, T) = 1 - \frac{\max(0, \cos(S, T))}{\max(|R_S|, |R_S|)} \quad \text{where} \quad d_s(S, T) \in [0, 1] \tag{8}$$

In contrast with semantic distance, in general, shape distance is calculated in three-dimensional space through algorithms based on image or 3D entity data, such as segmentation (Minaee et al., 2021) and shape histogram signature (Tombari et al., 2010). However, applying this concept to textual entities presents challenges. Vocabularies themselves do not have a specific spatial form; they are an abstract form of linguistic expression, and different linguistic environments and business domains bring differences to the application of the same word. Nevertheless, text-based language models based on strong contextual understanding have promoted the possibility of this practice. Since LLMs are trained based on semantic co-occurrence, their embedding vectors usually also reflect the overall information of the context, which is typically the functional concepts, usage scenarios, and semantic similarity. However, this shape similarity relationship is implicitly hidden in the knowledge of LLMs and cannot be represented stably and efficiently. For analogy reasoning tasks, the requirement for shape similarity is not high; humans pay more attention to the approximate shape when using visual analogies to reason about objects (Hashemi and Tubau, 2024). Therefore, we propose a shape language descriptor system for constructing an approximate similarity, aiming to associate the shape similarity between objects through shape descriptions with the same or similar semantics. The system includes four dimensions: Perceptual Shape Categorization (PSC), Height-to-Width Ratio (HWR), Structure, and Curvature Similarity, where PSC is constructed based on Recognition-by-Components (RBC) Theory (Biederman, 1987), which is the main factor in this system. For this purpose, we have constructed a set consisting of 36 basic shapes called geons. The other three dimensions are additional supplements to the shape description. Here, we specify their weights as PSC=0.4, and the other three dimensions are all 0.2. HWR is a 5-category description of the overall proportion of objects, structure is a 4-category description of the segmented structure, and Curvature Similarity contains 3 curvature descriptions, which are shown in Appendix A. The shape language descriptors will be used as enhanced data for the RAG strategy, allowing LLMs to generate accurate descriptions for objects' shape alignment. Once the shape descriptors are determined, we use cosine distance to calculate the text similarity of each dimension between the concepts before and after analogy, represented as:

$$\mathrm{Sim}(S, T) = \sum_{i} w_i \cos(\mathrm{Embed}(f_d(S_i)), \mathrm{Embed}(f_d(T_i))) \tag{9}$$

where f_d represents the LLMs as a shape description function, Embed() represents the calculation of the word embedding, and w_i represents the weight of each dimension. Combining Eq. (7), the morphological distance function is finally represented as:

$$d_m(S,T) = 1 - \frac{\sum_i \max(0, \mathrm{Sim}(S_i, T_i))}{\max(|R_S|, |R_S|)} \quad \text{where} \quad d_m(S,T) \in [0,1] \tag{10}$$

According to cognitive science and visual analogy theory, far-distance semantic relationships mean that the target domain and the source domain are not conceptually directly related. This large-span analogy can avoid overly conventional or obvious associations, while approximate morphological relationships ensure that they can share a set of perceived affordances (Siegel, 2014; Basri et al., 1998). Even if the semantics are different, people can still identify potential uses from familiar shapes. This helps to reduce cognitive burden and makes cross-domain conceptual transfer easier. Therefore, we permit a creative target analogy to be as far as possible in semantics, and as close as possible in approximate morphology, compared to the source domain. To avoid extreme values, the comprehensive distance score will be represented by geometric mean optimization:

$$d_s(S,T) \cdot (1 - d_m(S,T)) \rightarrow \max \tag{11}$$

where the weights of the two distance functions are equal. The maximum distance target is achieved when $d_s \rightarrow 1$ and $d_m \rightarrow 0$, and the minimum distance target is achieved when $d_s \rightarrow 0$ or $d_m \rightarrow 1$.

3.3. Reuse of analogical knowledge

Although analogy can help explore potentially related objects, merely using analogy does not guarantee the effective promotion of creative design. The main reason is that analogy itself does not directly produce new ideas, but provides a way of thinking. Since people's thinking is often limited by existing experience and mental set (Kray et al., 2006), simply obtaining the objects before and after the analogy is not enough. The key challenge lies in how to use the knowledge in the analogical objects to generate new ideas. Inspired by Han's combinational creativity (Han et al., 2019b), we propose a combinational block that reorganizes analogy knowledge for creative conceptual design. This block receives three parameters, namely the source concept

S from the PSB solution space S_a, the problem space P, and the target concept along with the ontology (T, O). Since S has mature functions and feasibility, it will be used as the base concept (i.e. the Base). The concept T obtained through analogy reasoning can fit the perceptual affordances of the source domain concept in form, which can serve as an additive concept (i.e., the Additive) to enhance or improve the creativity of the Base through combination. The problem space P includes the design objectives, scenario, and target users. Then, the LLMs improve or redesign the initial scheme of the source analogy according to the properties and functions of the target analogy by combination, as shown in Eq. (12):

$$(g_D, g_I, g_E) = f_{comb}(S, T, O, P) \tag{12}$$

where f_{comb} represents the LLMs as a combination function. g_D represents the generated creative conceptual design combining the source analogy with the target analogy; g_I represents the generated inspiration description that describes what functional or morphological feature of the analogy the creative idea comes from; and g_E represents the generated explanation which helps the user understand how the target analogy contributes to the creative idea. For those target analogies belonging to natural objects, we use bio-inspired signifiers to stimulate LLMs to learn from the ecological properties, morphology, and functions of natural objects. For example, the source domain is a TV tower, the target domain is a cactus of a natural object, the scene in the problem space is architectural design, and the goal is sustainable construction. Through LLM, g_D is described as "self-sustaining Television tower". g_I can be "storage capability and resilience of a cactus". g_E can be "the television tower can incorporate a rainwater harvesting system inspired by how cacti store water in their bodies. The outer shell of the tower could be designed with micro-channel structures that guide rainwater into a storage reservoir, which could then be used for cooling systems, reducing energy consumption and increasing sustainability".

In summary, the proposed computational concept product design approach, by integrating the knowledge reasoning and reuse ability of LLMs and the structured analogical retrieval mechanism, not only systematically retrieves cross-domain knowledge but also efficiently transforms this knowledge into creative concepts. This development overcomes the gap between analogy and creative generation in traditional design methods.

4. Case study

4.1. ViMimic

ViMimic, an interactive tool based on the proposed computational conceptual product design approach, was developed and implemented following the principles described in Section 3. Fig. 2 shows an example of the process of using this tool to generate an extendable vase design. The creative conceptual design of vase is generated by first reasoning the target analogy from "vase" and then reusing analogy knowledge to form a new combination. We use GPT-4o (Version: gpt-4o-2024-08-06) as the system's backend LLM, representing a leading state-of-the-art LLM. To reduce the experimental bias caused by LLM, we use default parameters (temperature=1.0, top_p=1.0). It is worth noting that our approach is designed to be compatible with any LLM. However, we recommend selecting models listed in the Chatbot Arena benchmarking (Chiang et al., 2024) as generative drivers. These models have been rigorously evaluated and validated for their performance, ensuring reliability and effectiveness in our approach. To mitigate the negative influence of hallucination of LLMz when generating content, ViMimic supports human-AI cross-validation through card-based interaction. We built ViMimic as a web-based interactive system using React.js and Python Flask. This allows users to inspect and select the analogies and combinational designs they consider appropriate. For image references that supplement visual knowledge, as seen in Fig. 2 (b3), we integrated Google Custom Search API[1] to retrieve images from Google's dataset. The zero-shot prompting for the analogy mapping block and the combination block was adopted to return the JSON format in a structured output by OpenAI's recommendation [2], which are shown in Appendix B and Appendix C.

4.2. Participants

18 designers were recruited for the study (Females = 9, Males = 9, Mean_age = 24.5, SD_age = 3.29) with backgrounds in product and industrial design (11 undergraduate background designers and seven master background designers). All participants were required to be familiar with innova-

[1] `https://developers.google.com/custom-search/v1/overview`

[2] `https://platform.openai.com/docs/guides/structured-outputs/introduction`

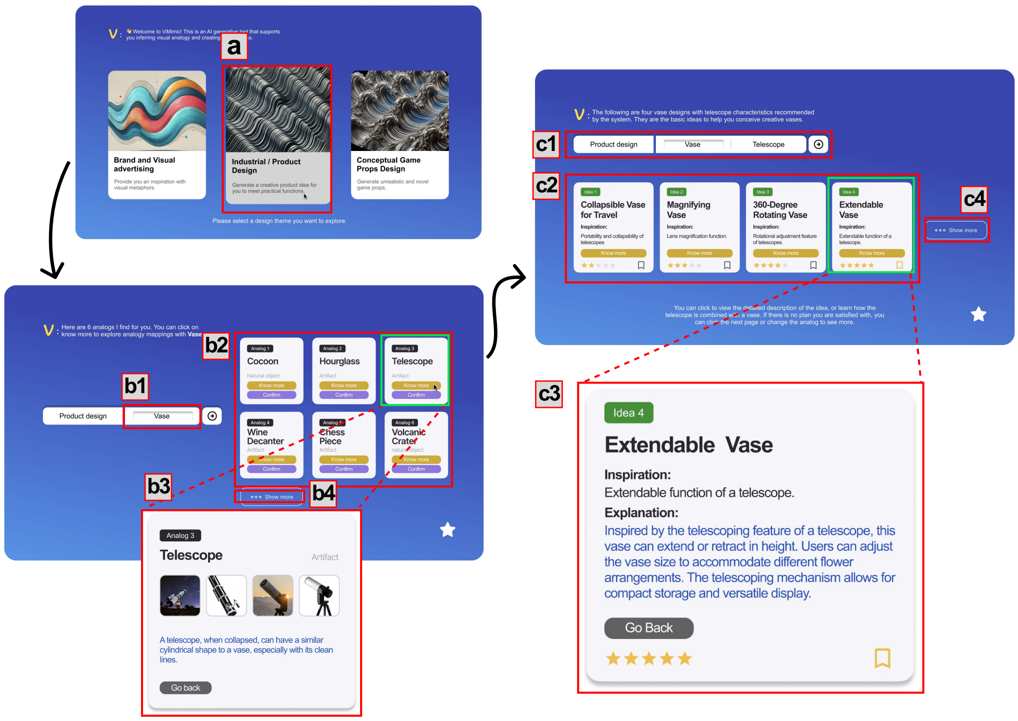

Fig. 2. The screenshot of the user interfaces of ViMimic. The usage process of the tool is from interfaces a to c: (a) selecting user scenario, (b1-b4) implementation of analogy mapping and scoring block, allowing users to select the generated analogy, (c1-c4) implementation of combination, allowing users to select the generated design idea.

tion design (Mean_design_experience = 4.07 years, SD_design_experience = 1.77) to mitigate the bias of the study.

4.3. Experiment and procedure

A within-subject study was conducted to compare and evaluate creative performance when using the tools based on our approach (i.e., ViMimic) versus current typical search and inspiration tools (i.e., the combined use of Google Search, Pinterest, and Behance) as a baseline. It is worth noting that our comparison is reasonable because these typical search and inspiration tools had integrated large language models and allowed AI-enabled search (e.g., Google integrated Gemini (Team et al., 2024) into its search engine before the experiment was implemented.

The participants were asked to use the assigned tools to complete a combinational design by analogy. Three design topics were provided: 1) kitchen knife, 2) umbrella, and 3) hair dryer. The reason for choosing these three concepts is that they are typical everyday product designs. These familiar objects invoke users' historical perceptions. It can be challenging to give these objects a new form of creativity than to design a brand-new creative concept that is unknown to us. Considering this, the three components in the problem space are all regarded as ordinary users in general scenarios, and therefore corresponding parameters are not specified in ViMimic.

Before the experiment began, participants received a 5-minute introduction to visual analogy. During this, we explained the definition of visual analogy, how to reason using it, and how to generate creative design ideas based on visual analogy through a combinational approach. To help participants better understand the concept, we provided several examples of visual analogy and design-by-analogy to offer a more intuitive understanding. We also clarified that the goal of the study is to verbally describe creative ideas rather than deliver a fully completed design. Therefore, the whole design does not need to be very detailed, just to produce a creative direction. The total duration of the study was 45 minutes. Participants were required to complete a 20-minute baseline design task and a 20-minute ViMimic-assisted design task. There was a 5-minute break between the two tasks. To avoid order effects, we randomly balanced the order of the two tasks and rotated the concepts within the same design context for the second task (e.g., P1T1: kitchen knife → P1T2: umbrella, P2T1: umbrella → P2T2: hair dryer, P3T1: hair dryer → P3T2: kitchen knife). During the tasks, participants were asked to record the divergent analogies. Fig. 3 shows some of the conceptual designs along with an inspirational analogy generated by ViMimic.

4.4. Evaluation

The goal of the evaluation is to investigate the overall creative performance using our approach, including expert evaluation, objective measures, and self-assessment.

The expert evaluation employs a 5-point Likert-scale Consensual Assessment Technique (CAT) (Amabile, 1982) to assess novelty, completeness, functionality, and flexibility of the generated conceptual designs. We recruited four experts with 8-10 years of creative design experience to rate the results generated by both the baseline and ViMimic. To avoid data sparsity issues when analyzing Fleiss' Kappa (Falotico and Quatto, 2015), we quantized the

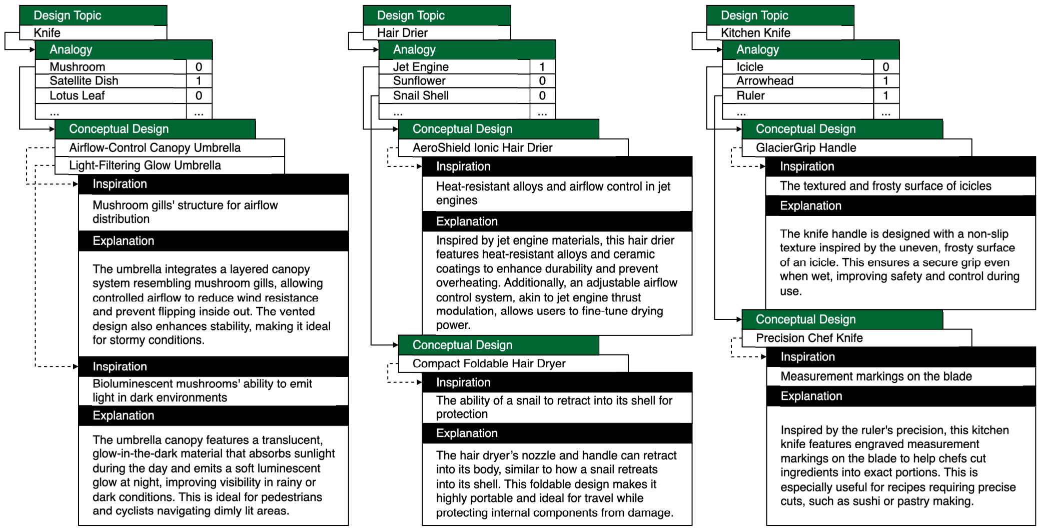

Fig. 3. Reports of the generated analogies and the conceptual designs by ViMimic using our approach, where 0 represents natural objects and 1 represents artifacts.

data for each evaluation dimension (combining scores of 1-2 as "Low," 3 as "Medium," and 4-5 as "High").

The objective measure is formed by the combination of diversity and efficiency. The time taken for completing analogy mapping and conceptual design combination, and the number of identified source analogs participants documented were recorded. We constructed two convex hulls (Avis and Bremner, 1995) containing 36-point sets, representing the number of ideas generated and the time taken in both the ViMimic and Baseline phases. The orientations of the centroids of these convex hulls were used to assess overall support of our approach for divergent thinking and creative efficiency.

The self-assessment consisted of a questionnaire survey, NASA-TLX, and an interview, aiming to evaluate creative experiences and cognitive load from designers. A survey using a 5-point Likert scale (Joshi et al., 2015) was conducted to gather feedback from participants in three areas: visual analogy mapping (Q1-4), creative design-by-visual analogy (Q5-8), and creativity stimulation (Q9-11). After the survey, each participant was asked to complete a NASA-TLX (Hart, 2006) questionnaire to measure their perceived cognitive load during the ideation process. Participants were then subjected to a semi-structured interview to delve deeper into how our approach supports their creativity.

4.5. Results

4.5.1. Expert evaluation

The results of CAT are illustrated in Table 1, which show a high creativity performance in generating conceptual product design. The results indicated that ViMimic-generated conceptual designs were superior to the baseline. Compared to the baseline, the overall creativity of the conceptual design generated by ViMimic has increased by 43.2%. Experts agreed that the conceptual designs of ViMimic are more novel (Novelty_Mean_diff = 0.76, moderate agreement) and complete (Completeness_Mean_diff = 0.82, substantial agreement) compared to the baseline. Although there is no significant agreement in the areas of functionality and flexibility, the CAT scores still indicated that ViMimic had a stronger ability to support creative output.

Table 1: The CAT results were quantized from the original 5-point Likert scale into a range of 1-3. Scores of 1-2 were categorized as 'Low (1)', 3 as 'Medium (2)', and 4-5 as 'High (3)'. P-values were interpreted as *: p<0.05; **: p<0.01; ***: p<0.001. Fleiss' Kappas were interpreted as follows: <0.33: no agreement; 0.33 < Kappa < 0.66: moderate agreement; >0.66: substantial agreement.

		Mean (P1-P18)	SD (P1-P18)	Fleiss' Kappa	Cohen's d	Sig.
Novelty	ViMimic	2.25	0.55	0.48	1.29	***
	Baseline	1.48	0.62	0.72		
Completeness	ViMimic	2.29	0.63	0.79	1.37	***
	Baseline	1.47	0.55	0.52		
Functionality	ViMimic	2.00	0.71	0.44	0.57	**
	Baseline	1.63	0.53	0.17		
Flexibility	ViMimic	1.87	0.69	0.22	0.98	**
	Baseline	1.27	0.51	0.53		

4.5.2. Objective measure

The results of the convex hull analysis are shown in Fig. 4, representing the number of ideas generated and the time taken in both the ViMimic and Baseline phases. In the case that the area of the two convex hulls is not significantly different, the closer the data points are to the left of the X-axis, the less time participants took to complete tasks, indicating higher efficiency.

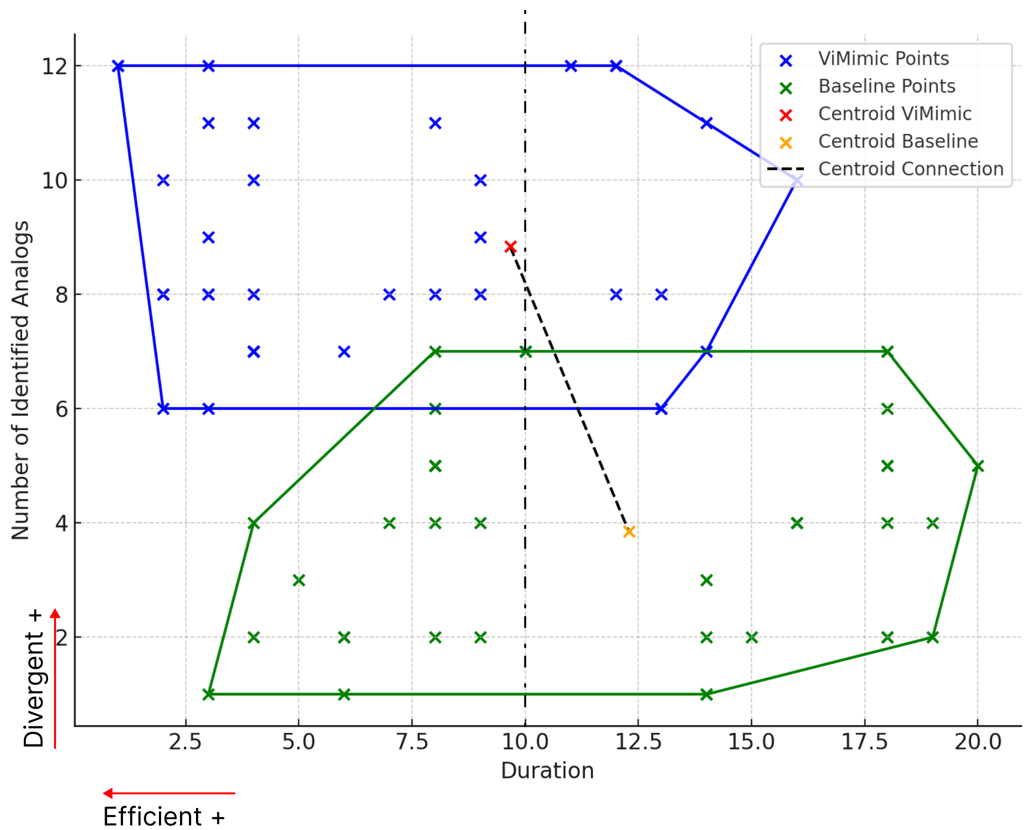

Fig. 4. Convex hulls of duration / number of identified analogies in both ViMimic and baseline.

The closer the data points are to the top of the Y-axis, the more analogies participants generated or found interesting. The convex hull area of ViMimic was only 9.03% smaller than the baseline (Convex_hull_area_ViMimic = 77.5 units^2, Convex_hull_area_Baseline = 84.5 units^2). ViMimic's overall area and centroid were positioned above and to the left of the baseline (ViMimic centroid = (9.67, 8.83), Baseline centroid = (12.29, 3.86)). To quantify the overall improvement in the distribution, we calculated the Euclidean distance from each method's centroid to an ideal reference point (x=0, y=15). The corresponding distances to the ideal point were 11.47 and 16.60, respectively. This represents a 30.9% reduction in distance, indicating that ViMimic outperformed the baseline in supporting designers' divergent visual analogies and creative design generation.

4.5.3. Self-assessment

A t-test was conducted to analyze the questionnaire survey. As shown in Fig. 5, the results show that participants were more satisfied with the efficiency of using ViMimic to obtain visual analogies (Q1, Mean_diff = 1.55, $p < 0.001$). They also expressed greater satisfaction with the variety of cross-domain analogies provided by ViMimic (Q3, Mean_diff = 2.39, $p < 0.001$). They felt that the conceptual design ideas were more unique and innovative than usual (Q5, Mean_diff = 1.16, $p < 0.001$, ViMimic_SD = 0.81, Baseline_SD = 0.87), and the generated concepts were relevant to the scenario's requirements (Q6, Mean_diff = 0.72, $p < 0.05$, ViMimic_SD = 0.97, Baseline_SD = 0.82). The results also suggest that ViMimic helped participants better understand analogy-inspired conceptual designs (Q8, Mean_diff = 1.28, $p < 0.001$, ViMimic_SD = 0.64, Baseline_SD = 0.87) and generate a variety of creative designs (Q10, Mean_diff = 1.33, $p < 0.001$; Q11, Mean_diff = 1.94, $p < 0.001$). However, there was no significant difference between the baseline and ViMimic in terms of analogy decision-making (Q2, Mean_diff = 0.72, $p > 0.05$, ViMimic_SD = 1.14, Baseline_SD = 0.92; Q4, Mean_diff = 0.28, $p > 0.05$, ViMimic_SD = 0.99, Baseline_SD = 0.90).

The results of the NASA-TLX are presented in Table 2. The mental demand of using ViMimic for analogy was significantly lower than that in the baseline condition (Mean_Vi_md = 2.72, Mean_Ba_md = 4.66, $p < 0.001$). Designers' mental and physical demands are reduced by 41.6% and 29.2% respectively. In contrast to the baseline, ViMimic performed a large amount of knowledge retrieval and computation for users, reducing their physical effort (Mean_Vi_pd = 1.33, Mean_Ba_pd = 1.88, $p < 0.05$; Mean_Vi_fs = 2.66, Mean_Ba_fs = 3.83, $p < 0.05$). However, the result of the effort suggests that they could not determine which analogy would be more beneficial for the subsequent creative combinational design. More than half of the participants indicated that, since the analogy objects were all familiar and common, it was difficult to judge which analogy would be better for later creativity. They found that deciding on a good analogy did not feel significantly easier compared to the baseline (Mean_Vi_ef = 3.83, Mean_Ba_ef = 4.83). Despite this, participants expressed a preference for continuing to use ViMimic, as it allowed them to quickly access analogies and reduced cognitive effort, which self-thinking or using traditional AI tools could not achieve. Overall, ViMimic is highly effective in supporting divergent visual analogy generation. Users experienced reduced cognitive load and gained access to a

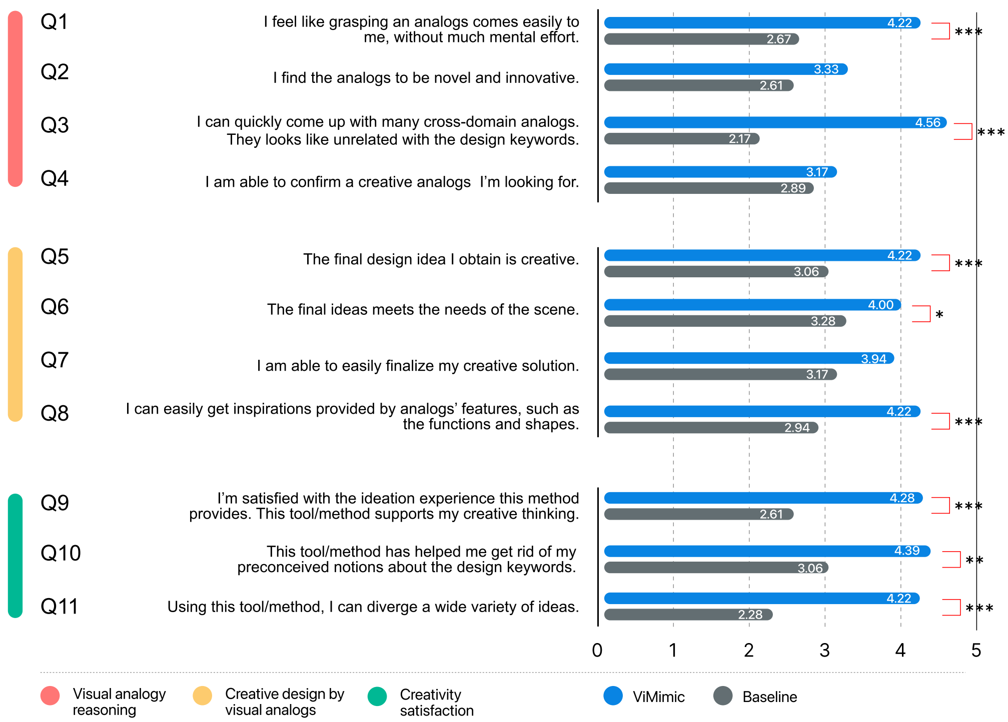

Fig. 5. The questionnaire asked participants about three dimensions, including visual analogical reasoning, creative design-by-visual analogy, and creativity satisfaction on a 5-point Likert scale. The results were analyzed using significance levels: *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

variety of analogy options. However, the system did not achieve high satisfaction in aiding users with analogy decision-making. In the results of the semi-structured interview, P11 noted, *"Compared to the baseline, I had no idea there would be so many interesting creative combinations. I felt like my creative thinking was opened up"*. Moreover, ViMimic helped participants better understand analogy-inspired creative ideas. P17 explained, *"I could quickly build on the analogies' foundation and add new ideas, which really stimulated my creativity. For example, an Airflow-Control Canopy Umbrella is inspired by a mushroom gills' structure, which integrates a layered canopy system resembling mushroom gills, allowing controlled airflow to reduce wind resistance and prevent flipping inside out. The vented design also enhances stability, making it ideal for stormy conditions. If ViMimic had not presented these potential ideas, it might have taken me much longer to come up with it"*. In the baseline, half of the users found it difficult to think of an interesting analog. P10 mentioned, *"In the baseline, I would transition from thinking with my brain to using computer tools, and they consumed a lot of effort. Since many results were irrelevant, I had to keep searching and trying to find analogies"*.

Table 2: NASA-TLX questionnaire results comparing ViMimic and Baseline on a 5-Likert Scale. n.s.: $p > 0.05$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

	ViMimic		Baseline		p-value	Sig.
	Mean	SD	Mean	SD		
Mental Demand (md)	2.72	0.57	4.66	1.81	0.000	***
Physical Demand (pd)	1.33	0.59	1.88	0.96	0.037	*
Temporal Demand (td)	2.94	1.51	3.16	1.42	0.663	*n.s.*
Performance (pf)	4.55	0.98	4.05	1.89	0.207	*n.s.*
Effort (ef)	3.83	1.20	4.83	1.42	0.031	*
Frustration (fs)	2.66	1.08	3.83	2.23	0.042	*

5. Discussion

5.1. Discussion of methodology

5.1.1. Enhancing analogical reasoning by LLMs

Human designers often struggle with analogical thinking and generating concepts due to inherent cognitive limitations and bounded knowledge spaces (Lieder and Griffiths, 2020). Manually identifying relevant analogies across domains requires extensive experience and expertise. However, even skilled designers may overlook distant yet valuable inspirations due to cognitive biases and knowledge constraints (Crilly, 2015). The complexity of processing and integrating diverse knowledge makes it difficult to innovatively explore a broad solution space. In contrast, our approach leverages the capabilities of LLMs to perform analogical reasoning and generate innovative conceptual design solutions. Through case studies, the experimental results further support the feasibility and effectiveness of the proposed approach, albeit with limited testing. The approach presented here does not rely on external domain-specific knowledge bases. By leveraging the vast, implicit understanding of patterns, principles, and structures learned from extensive pretraining data, our approach can autonomously discover diverse analogies and synthesize novel combinational design concepts.

One of the major challenges in employing large language models (LLMs) for reasoning tasks is that the entire generation process functions as a black-box (Sun et al., 2022). This implies that even with techniques such as fine-tuning and prompt engineering, the generated outputs may still be incorrect (Pan et al., 2023). A common solution is to have human reviewers verify the generated content to assess its accuracy and reliability. However, this kind of solution requires users to possess domain-specific knowledge relevant to the task. For evaluating analogies, our approach introduces an automated computational scoring strategy, making the reasoning process of LLMs more transparent and controllable while reducing reliance on human verification.

Moreover, the approach presented demonstrates the potential capability of mental imagery (Kosslyn et al., 2006) recognition by leveraging LLMs. Unlike traditional machine vision systems that primarily focus on feature extraction and classification, the shape language descriptor system can represent cognitive shape properties of an object using natural language. This capability highlights the potential of LLMs to bridge the gap between perceptual recognition and conceptual abstraction (Yee, 2019). This not only advances LLM-driven design but also sheds light on how PDMS can approximate human-like cognitive processes in analogical reasoning tasks.

5.1.2. Reusing knowledge for creative discovery

Although analogy serves as a valuable ideation method, its outcomes often remain within the realm of common knowledge, thereby imposing constraints on innovation (Champagne, 2015). As a result, designers may find their ability to associate with these familiar concepts inhibited. The approach presented here reuses analogical knowledge to address this limitation by retrieving connections between source and target domains based on combinational creativity method. The results of the case study reveal that individuals can discover new solutions by linking seemingly unrelated things. This combinational approach is well-suited for transitioning from visual analogy to creative design. However, the difficulty of recombining common knowledge depends on an individual's ability to understand and manipulate it (Nusbaum and Silvia, 2011). The proposed approach can assist designers by searching for combinational inspirations and explaining why such combinations work, thus easing the cognitive load in transitioning from analogy to design. The approach ensures that creative discoveries remain relatable, making them more practical and applicable in real-world design contexts.

While much research has focused on training generative models to produce novel outcomes (Chen et al., 2019; Edwards et al., 2024), we argue that instead of forcing machines to generate entirely new, unheard-of knowledge, it is more productive to have them reliably generate human common knowledge (e.g., through analogy) and help users draw creative inspiration from it. In other words, the analogical mappings provided by our approach as an information retrieval mechanism, similar to existing heuristic techniques (e.g., brainstorming (Rawlinson, 2017) and reverse thinking (Myerson, 2016)) and knowledge graph tools (e.g., ASKnatureNet (Chen et al., 2024a) and WikiLink (Zuo et al., 2022)). While the retrieved concepts stem from familiar human knowledge, the challenge lies in ensuring that this knowledge can be applied to keyword-driven design tasks. Our results show that ViMimic helps designers understand the relevance of visual analogy from other domains to the nature of the problem through inspiration and explanation. We believe the reason for this relevance is that two visually analogical objects often share the same set of perceived affordances (Norman, 2004), which allows them to be applied to similar functional problems. For example, while a "ruler" provides a measurement function, it does not inherently possess the cutting properties of a "kitchen knife". However, if designers break free from the conventional use of the ruler, its shape could potentially support cutting tasks. Since affordances are perceived visually, we suggest that when the problem space of a design task is not explicitly defined, this functional interchangeability can be used to reverse-explore concepts. For instance, a multifunctional ruler could be designed to cut, with its cutting property derived from the affordance of the ruler's blade-like shape, akin to a knife's thin edge.

5.1.3. Generating conceptual design for product innovation

The findings of the case study indicate that the approach presented can, to some extent, mitigate anchoring effects (Furnham and Boo, 2011), though this impact varies depending on the specific context and individual characteristics. Anchoring effects were more pronounced in the baseline condition. When participants engaged in analogical reasoning, they remained influenced by initial anchors. For example, P14 stated, *"I associated a handgun with a hair dryer, but I didn't know how to combine them. Their connection was too weak"*. This may be because participants organizing their own visual analogies might introduce many factors related to their current domain. While the handgun visually resembles the hair dryer, they have no functional relationship. In contrast, P14 was interested in the solution "Quick-release detachable air nozzle: This hair dryer features a detachable nozzle that can be quickly swapped out, inspired by the fast and efficient magazine change of a handgun. It allows users to switch between different styling nozzles for varied airflow with ease" in ViMimic, and explained in the interview, *"I found this idea very creative. Initially, I kept thinking about the handgun's firing function and was influenced by the hair dryer's airflow function. However, I overlooked the handgun's magazine change system. This idea opened up new thinking for me"*.

Another finding is that the approach presented can help alleviate the issue of design fixation caused by generative AI. Previous empirical studies have shown that designers are easily influenced by AI-generated images (Wadinambiarachchi et al., 2024). They tend to adopt the image solutions provided by AI and overlook the exploration of more novel and unique design paths. To address the design fixation problem introduced by generative AI, more abstract textual descriptions are used as output. These solutions were meant to represent rough ideas rather than fully detailed designs for a designer to further refine.

5.1.4. Implications for product design and manufacturing systems

This approach primarily addresses creativity aspects in conceptual design, focusing on enhancing ideation through analogy-based reasoning with the support of LLMs. However, its direct applicability to engineering and manufacturing remains limited at this stage. To fully support PDMS, further integration of engineering and manufacturing knowledge is necessary.

This includes incorporating domain-specific constraints, materials knowledge, process feasibility, and other practical considerations inherent to real-world product development.

Furthermore, existing data, methods, and technologies from the PDMS and engineering informatics communities can be structured and formalized as knowledge bases or workflows to complement and enhance the proposed approach. By integrating structured engineering knowledge such as reverse engineering (Raja and Fernandes, 2007) and function–behavior–structure models (Gero and Kannengiesser, 2004), the approach could evolve into a more comprehensive design and engineering support system. We believe that while this approach introduces a novel perspective on analogy-driven conceptual product design, it can also benefit from the well-established practices and structured knowledge representations within PDMS research and practice.

While the experimental validation was conducted within the scope of product design cases, we suggest that the approach can be adapted to early-stage creative information retrieval processes in various design engineering domains. The analogy retrieval guided by LLMs can be extended to other fields such as mechanical system design, civil engineering and biomedical device development.

5.2. Limitations and future work

Currently, the approach presented offers limited decision-making support. Although it reduces the cognitive load associated with generating analogies, it does not significantly assist designers in discriminating between these analogies. It is difficult to select the best analogy or creative idea from various options, as all can seem equally viable without clear direction. Future research and development could include enhanced decision-support features, such as algorithms that highlight or rank analogies based on designers' preferences or specific design contexts, learning from interactions, and prioritizing analogies with higher creative potential. Despite the advantages of the scoring block in our approach, the success of this strategy depends on several critical factors, including the definition of analogy relationship, computational distances, and the enhancement of its generalization capability, which could be inflexible when evaluating different analogy methods. These limitations require further in-depth research and exploration in the future.

This research focuses on language-based methods to identify novel ideas in conceptual product design. However, we recognize that some design issues are often complex, ambiguous, and not easily captured through verbal or textual descriptions alone. Integrating visual representations into the problem-finding framework could enrich the identification and articulation of user requirements and support designers in capturing aspects that may be difficult to verbalize (Dahl et al., 2001). Although this integration is beyond the scope of the current study, we see it as a promising direction for future research, particularly in expanding the applicability of our framework to more visually-driven design domains.

While LLMs offer broad analogy-based reasoning and combination at the conceptualization stage, integrating expert technical knowledge is crucial for new product design and development (Wang et al., 2024). Future work on the implementation stage in product design could incorporate private or domain-specific datasets to fine-tune or augment the LLM, or adopt a human-in-the-loop setup to ensure technical accuracy.

6. Conclusions

This paper introduces a novel computational conceptual product design methodology that integrates large language models with an analogy-based structured retrieval mechanism to support and enhance creativity in the early stages of conceptual product design. By leveraging the extensive world knowledge and contextual understanding capabilities of LLMs, the proposed approach enables more effective analogy reasoning, concept combination, and knowledge reuse, all while maintaining transparency and controllability throughout the design process.

The key contributions of this work are threefold. First, it demonstrates how LLMs can function as artificial analogy agents capable of generating meaningful and diverse analogies that inspire innovative design ideas. Secondly, a structured retrieval mechanism that guides the reasoning process, ensuring relevance and coherence while reducing the cognitive burden on human designers is presented. Thirdly, through comparative experimental validation, the approach is shown to significantly enhance design outcomes in terms of novelty, diversity, and reasoning efficiency, emphasizing its potential as a practical tool for design ideation.

Overall, this approach highlights the potential of LLM-based computational analogy in facilitating design innovation through knowledge transformation. By bridging AI techniques with cognitive principles, the proposed approach provides a foundation for more intelligent, interpretable, and creativity-enhancing product design systems.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research is supported by National Key R&D Program of China (2022YFB3303304).

Appendix A. Shape Language Descriptor

Table A.1: Pre-defined 36 Perceptual Shapes Categorization

Number	Geon	Example Object
1	Cylinder	Soda can
2	Cone	Traffic cone
3	Sphere	Basketball
4	Torus	Inner tube (donut float)
5	Brick (Rectangular Prism)	Shoe box
6	Pyramid	Egyptian pyramid
7	Wedge	Door wedge
8	Barrel (Curved Cylinder)	Beer barrel
9	Disc (Flat Cylinder)	Coin
10	Ellipsoid	Rugby ball
11	Arch	Bridge arch
12	Handle (Bent Tube)	Mug handle
13	Tapered Cylinder	Drumstick
14	Curved Wedge	Boomerang
15	Trapezoidal Prism	Trapezoidal roof
16	Double Cone	Hourglass
17	Half Cylinder	Half-pipe
18	Horseshoe	Horseshoe
19	Crescent	Crescent moon
20	U-Shaped Tube	Staple remover
21	Y-Shaped Structure	Slingshot

22	L-Shape	Bookend
23	T-Shape	Hammer head
24	Cross-Shaped Prism	Medical cross sign
25	Hourglass Shape	Egg timer
26	Flat Slab	Chopping board
27	Triangular Prism	Toblerone chocolate
28	Pipe	Water pipe
29	Box with Rounded Corners	Rounded jewelry box
30	Elliptical Cylinder	Vitamin pill
31	Irregular Lobe	Kidney bean
32	Curved Plate	Folded sheet metal
33	Hollow Tube	PVC pipe
34	Forked Branch Shape	Tree branch fork
35	Parallel-Piped	Shipping container
36	Handle	Suitcase handle

Table A.2: Height-to-Width Ratio

Number	HWR	Example Object
1	Square-like	Square, cube
2	Moderately Elongated	Books, bricks
3	Highly Elongated	Pen, remote control
4	Extremely Elongated	Bamboo pole, long sword
5	Flat	Plates, leaves

Table A.3: Segmented Structure

Number	Structure	Example Object
1	Single Component	Leaves, glass balls
2	Segmented	Small knife (blade + handle), TV tower
3	Branched	Cactus, branches
4	Complex Composite	Chair (backrest + seat + legs)

Table A.4: Curvature Similarity

Number	Curvature	Example Object
1	Flat or Slightly Curved	Books, cutting boards
2	Moderately Curved	Leaves, flat ovals
3	Highly Curved	Sphere, donut

Appendix B. Prompting For Analogy Reasoning

System_prompt:

```
You are an assistant who is proficient in visual analogy. You
   need to follow the following requirements to infer concepts
   : that is, the output concept and the input concept are
   similar in approximate shape distance, but are far in
   semantic distance. Please infer 10 suggested objects
   according to user input, and return them as a valid JSON
   object in the following format:
{
  "Concepts": [
    {
      "Concept": "<string>", // it can be any artifacts or
          natural objects.
      "Category": "<string>", // If the object obtained by
          reasoning is an artifact, please mark it as 1. If it
          is a natural object, please mark it as 0.
      "Explanation": "<string>" // Only explain how the object
          is inspired visually by the input.
    },
    {
      "Concept": "<string>",
      "category": "<string>",
      "explanation": "<string>"
    }
    ...
  ]
}
```

User_prompt:

```
Please enter a design keyword you want to reason by analogy (
    for example: chair)
Designer input: [INPUT0]
```

Appendix C. Prompting For Creative Combination

System_prompt:

```
You are an expert designer. You know how to use visual analogy
     skillfully, and you can use analogied concepts for
    creative inspiration. You need to refer to the concept of
    the target domain to propose a creative design scheme for
    the design of the source domain, and describe how the
    function or appearance characteristics of the target domain
     can help you inspire the creative design of the source
    domain concept.

Your output will be different according to the user's scene
    needs: in product and industrial design scene, your output
    will focus on actual functions, production, and cost; in
    brand and advertising scene, your output will focus on
    obtaining metaphor from the target domain concept to design
     a poster or advertising by a creative way; in game and
    architectural scene, your output will focus on high novelty
     or unrealistic creative design. Return your output as a
    valid JSON object in the following response format:
{
    "Solutions": [ // list four suggested design solutions of
        the source domain concept, inspired by the features of
        the target domain concept.
    {
      "Design": "<string>", // A product in product design, or
          a slogan in brand and advertising, or a name in game
          and architecture.
      "Inspiration": "<string>", // What feature is inspired by
           the concept of the target domain?
      "Explanation": "<string>" // Explain how the design is.
    }
  ]
}
```

User_prompt:

```
Please select an analog you want to use as a creative element
    for the [INPUT0].

Source concept [INPUT0]
Target concept [INPUT1]
Scene: [INPUT2]
```

References

Altshuller, G.S., 1984. Creativity as an exact science. crc Press.

Amabile, T.M., 1982. Social psychology of creativity: A consensual assessment technique. Journal of personality and social psychology 43, 997.

Avis, D., Bremner, D., 1995. How good are convex hull algorithms?, in: Proceedings of the eleventh annual symposium on Computational geometry, pp. 20–28.

Baby, M., Nellippallil, A.B., 2024. An information-decision framework for the multilevel co-design of products, materials, and manufacturing processes. Advanced Engineering Informatics 59, 102271.

Basri, R., Costa, L., Geiger, D., Jacobs, D., 1998. Determining the similarity of deformable shapes. Vision Research 38, 2365–2385.

Bettig, B., Bapat, V., Bharadwaj, B., 2005. Limitations of parametric operators for supporting systematic design, in: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, pp. 131–142.

Bian, N., Han, X., Sun, L., Lin, H., Lu, Y., He, B., Jiang, S., Dong, B., 2024. Chatgpt is a knowledgeable but inexperienced solver: An investigation of commonsense problem in large language models, in: LREC/COLING, pp. 3098–3110.

Biederman, I., 1987. Recognition-by-components: a theory of human image understanding. Psychological review 94, 115.

Blanchette, I., Dunbar, K., 2000. How analogies are generated: The roles of structural and superficial similarity. Memory & cognition 28, 108–124.

Boden, M.A., 2009. Computer models of creativity. Ai Magazine 30, 23–23.

Boothroyd, G., Dewhurst, P., Knight, W.A., 2010. Product design for manufacture and assembly. CRC press.

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.

Champagne, M., 2015. Experience and life as ever-present constraints on knowledge. Metaphilosophy 46, 235–245.

Chandrasegaran, S.K., Ramani, K., Sriram, R.D., Horváth, I., Bernard, A., Harik, R.F., Gao, W., 2013. The evolution, challenges, and future of knowledge representation in product design systems. Computer-aided design 45, 204–228.

Chen, L., Cai, Z., Jiang, Z., Luo, J., Sun, L., Childs, P., Zuo, H., 2024a. Asknaturenet: A divergent thinking tool based on bio-inspired design knowledge. Advanced Engineering Informatics 62, 102593.

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

Chen, L., Wang, P., Dong, H., Shi, F., Han, J., Guo, Y., Childs, P.R., Xiao, J., Wu, C., 2019. An artificial intelligence based data-driven approach for design ideation. Journal of Visual Communication and Image Representation 61, 10–22.

Chiang, W.L., Zheng, L., Sheng, Y., Angelopoulos, A.N., Li, T., Li, D., Zhu, B., Zhang, H., Jordan, M.I., Gonzalez, J.E., Stoica, I., 2024. Chatbot arena: an open platform for evaluating llms by human preference, in: ICML'24, JMLR.org. p. 30.

Childs, P., Han, J., Chen, L., Jiang, P., Wang, P., Park, D., Yin, Y., Dieckmann, E., Vilanova, I., 2022. The creativity diamond—a framework to aid creativity. Journal of Intelligence 10, 73.

Crilly, N., 2015. Fixation and creativity in concept development: The attitudes and practices of expert designers. Design studies 38, 54–91.

Dahl, D.W., Chattopadhyay, A., Gorn, G.J., 2001. The importance of visualisation in concept design. Design studies 22, 5–26.

Dathathri, S., Madotto, A., Lan, J., Hung, J., Frank, E., Molino, P., Yosinski, J., Liu, R., 2019. Plug and play language models: A simple approach to controlled text generation. arXiv preprint arXiv:1912.02164 .

Daugherty, J., Mentzer, N., 2008. Analogical reasoning in the engineering design process and technology education applications. Journal of Technology Education doi:10.21061/jte.v19i2.a.1.

Duan, Y., Zhang, J., 2022. A novel ai-based visual stimuli generation approach for environment concept design. Computational Intelligence and Neuroscience 2022, 8015492.

Edwards, K.M., Man, B., Ahmed, F., 2024. Sketch2prototype: rapid conceptual design exploration and prototyping with generative ai. Proceedings of the Design Society 4, 1989–1998.

Falotico, R., Quatto, P., 2015. Fleiss' kappa statistic without paradoxes. Quality & Quantity 49, 463–470.

Forbus, K.D., Gentner, D., Markman, A.B., Ferguson, R.W., 1998. Analogy just looks like high level perception: Why a domain-general approach to analogical mapping is right. Journal of Experimental & Theoretical Artificial Intelligence 10, 231–257.

Fu, K., Chan, J., Cagan, J., Kotovsky, K., Schunn, C., Wood, K., 2013. The meaning of "near" and "far": the impact of structuring design databases and the effect of distance of analogy on design output. Journal of Mechanical Design 135, 021007.

Furnham, A., Boo, H.C., 2011. A literature review of the anchoring effect. The journal of socio-economics 40, 35–42.

Gassmann, O., Zeschky, M., 2008. Opening up the solution space: The role of analogical thinking for breakthrough product innovation. Creativity and Innovation Management 17, 97–106.

Gentner, D., Holyoak, K.J., 1997. Reasoning and learning by analogy: Introduction. American psychologist 52, 32.

Gentner, D., Markman, A.B., 1997. Structure mapping in analogy and similarity. American psychologist 52, 45.

Gero, J.S., Kannengiesser, U., 2004. The situated function–behaviour–structure framework. Design studies 25, 373–391.

Gick, M.L., Holyoak, K.J., 1980. Analogical problem solving. Cognitive psychology 12, 306–355.

Goel, A.K., 1997. Design, analogy, and creativity. IEEE expert 12, 62–70.

Goel, A.K., Bhatta, S.R., 2004. Use of design patterns in analogy-based design. Advanced Engineering Informatics 18, 85–94.

Gorti, S.R., Gupta, A., Kim, G.J., Sriram, R.D., Wong, A., 1998. An object-oriented representation for product and design processes. Computer-aided design 30, 489–501.

Han, J., Hua, M., Shi, F., Childs, P.R., 2019a. A further exploration of the three driven approaches to combinational creativity, in: Proceedings of the Design Society: International Conference on Engineering Design, Cambridge University Press. pp. 2735–2744.

Han, J., Park, D., Shi, F., Chen, L., Hua, M., Childs, P.R., 2019b. Three driven approaches to combinational creativity: Problem-, similarity-and inspiration-driven. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 233, 373–384.

Hart, S.G., 2006. Nasa-task load index (nasa-tlx); 20 years later, in: Proceedings of the human factors and ergonomics society annual meeting, Sage publications Sage CA: Los Angeles, CA. pp. 904–908.

Hashemi, A., Tubau, E., 2024. Global relations versus object relations in visual analogies. Thinking & Reasoning , 1–27.

Holyoak, K.J., Thagard, P., 1996. Mental leaps: Analogy in creative thought. MIT press.

Hoque, A.M., Halder, P.K., Parvez, M.S., Szecsi, T., 2013. Integrated manufacturing features and design-for-manufacture guidelines for reducing product cost under cad/cam environment. Computers & Industrial Engineering 66, 988–1003.

Howard, T.J., Culley, S.J., Dekoninck, E., 2008. Describing the creative design process by the integration of engineering design and cognitive psychology literature. Design studies 29, 160–180.

Johnson, J., Hurst, A., Safayeni, F., 2023. Managing data-driven design: A survey of the literature and future directions. Proceedings of the Design Society 3, 2525–2534.

Joosten, J., Bilgram, V., Hahn, A., Totzek, D., 2024. Comparing the ideation quality of humans with generative artificial intelligence. IEEE Engineering Management Review 52, 153–164.

Joshi, A., Kale, S., Chandel, S., Pal, D.K., 2015. Likert scale: Explored and explained. British journal of applied science & technology 7, 396.

Koren, Y., Gu, X., Guo, W., 2018. Reconfigurable manufacturing systems: Principles, design, and future trends. Frontiers of Mechanical Engineering 13, 121–136.

Kosslyn, S.M., Thompson, W.L., Ganis, G., 2006. The case for mental imagery. Oxford University Press.

Kray, L.J., Galinsky, A.D., Wong, E.M., 2006. Thinking within the box: The relational processing style elicited by counterfactual mind-sets. Journal of personality and social psychology 91, 33.

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.

Lee, S., Sim, W., Shin, D., Seo, W., Park, J., Lee, S., Hwang, S., Kim, S., Kim, S., 2024. Reasoning abilities of large language models: In-depth analysis on the abstraction and reasoning corpus. ACM Transactions on Intelligent Systems and Technology .

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.t., Rocktäschel, T., et al., 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in neural information processing systems 33, 9459–9474.

Lieder, F., Griffiths, T.L., 2020. Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. Behavioral and brain sciences 43, e1.

Ma, K., Grandi, D., McComb, C., Goucher-Lambert, K., 2023. Conceptual design generation using large language models, in: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers. p. V006T06A021.

Marvin, G., Hellen, N., Jjingo, D., Nakatumba-Nabende, J., 2023. Prompt engineering in large language models, in: International conference on data intelligence and cognitive informatics, Springer. pp. 387–402.

Mascitelli, R., 2000. From experience: harnessing tacit knowledge to achieve breakthrough innovation. Journal of Product Innovation Management: an International Publication of the Product Development & Management Association 17, 179–193.

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., Terzopoulos, D., 2021. Image segmentation using deep learning: A survey. IEEE transactions on pattern analysis and machine intelligence 44, 3523–3542.

Myerson, J., 2016. Scaling down: Why designers need to reverse their thinking. She Ji: The Journal of Design, Economics, and Innovation 2, 288–299.

Nooralahzadeh, F., Bekoulis, G., Bjerva, J., Augenstein, I., 2020. Zero-shot cross-lingual transfer with meta learning. arXiv preprint arXiv:2003.02739 .

Norman, D., 2004. Affordances and design. Unpublished article, available online at: http://www. jnd. org/dn. mss/affordances-and-design. html , 14.

Nusbaum, E.C., Silvia, P.J., 2011. Are intelligence and creativity really so different?: Fluid intelligence, executive processes, and strategy use in divergent thinking. Intelligence 39, 36–45.

Pan, L., Saxon, M., Xu, W., Nathani, D., Wang, X., Wang, W.Y., 2023. Automatically correcting large language models: Surveying the landscape of diverse self-correction strategies. arXiv preprint arXiv:2308.03188 .

Perković, G., Drobnjak, A., Botički, I., 2024. Hallucinations in llms: Understanding and addressing challenges, in: 2024 47th MIPRO ICT and Electronics Convention (MIPRO), IEEE. pp. 2084–2088.

Peruzzini, M., Pellicciari, M., 2017. A framework to design a human-centred adaptive manufacturing system for aging workers. Advanced Engineering Informatics 33, 330–349.

Raja, V., Fernandes, K.J., 2007. Reverse engineering: an industrial perspective. Springer Science & Business Media.

Rawlinson, J.G., 2017. Creative thinking and brainstorming. Routledge.

Reynolds, L., McDonell, K., 2021. Prompt programming for large language models: Beyond the few-shot paradigm, in: Extended abstracts of the 2021 CHI conference on human factors in computing systems, pp. 1–7.

Serrat, O., Serrat, O., 2017. The scamper technique. Knowledge solutions: tools, methods, and approaches to drive organizational performance , 311–314.

Siegel, S., 2014. Affordances and the contents of perception. Does perception have content , 39–76.

Sun, T., Shao, Y., Qian, H., Huang, X., Qiu, X., 2022. Black-box tuning for language-model-as-a-service, in: International Conference on Machine Learning, PMLR. pp. 20841–20855.

Team, G., Mesnard, T., Hardin, C., Dadashi, R., Bhupatiraju, S., Pathak, S., Sifre, L., Rivière, M., Kale, M.S., Love, J., et al., 2024. Gemma: Open models based on gemini research and technology. arXiv preprint arXiv:2403.08295 .

Tombari, F., Salti, S., Di Stefano, L., 2010. Unique signatures of histograms for local surface description, in: Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part III 11, Springer. pp. 356–369.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Advances in neural information processing systems 30.

Vinichenko, M.V., Melnichuk, A.V., Karácsony, P., 2020. Technologies of improving the university efficiency by using artificial intelligence: Motivational aspect. Entrepreneurship and sustainability issues 7, 2696.

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 2024 CHI Conference on Human Factors in Computing Systems, pp. 1–18.

Wang, Z., Fu, C., Niu, S.f., Hu, S.j., 2024. Research on bio-inspired product design based on knowledge graph and semantic fusion diffusion model. Advanced Engineering Informatics 62, 102797.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q.V., Zhou, D., et al., 2022. Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems 35, 24824–24837.

Yaner, P.W., Goel, A.K., 2006. Visual analogy: Viewing analogical retrieval and mapping as constraint satisfaction problems. Applied Intelligence 25, 91–105.

Yang, Z., Gan, Z., Wang, J., Hu, X., Lu, Y., Liu, Z., Wang, L., 2022. An empirical study of gpt-3 for few-shot knowledge-based vqa, in: Proceedings of the AAAI conference on artificial intelligence, pp. 3081–3089.

Ye, X., Durrett, G., 2022. The unreliability of explanations in few-shot prompting for textual reasoning. Advances in neural information processing systems 35, 30378–30392.

Yee, E., 2019. Abstraction and concepts: when, how, where, what and why?

Yu, J., Wang, X., Tu, S., Cao, S., Zhang-Li, D., Lv, X., Peng, H., Yao, Z., Zhang, X., Li, H., et al., 2023a. Kola: Carefully benchmarking world knowledge of large language models. arXiv preprint arXiv:2306.09296 .

Yu, Y., Zhuang, Y., Zhang, J., Meng, Y., Ratner, A.J., Krishna, R., Shen, J., Zhang, C., 2023b. Large language model as attributed training data generator: A tale of diversity and bias. Advances in Neural Information Processing Systems 36, 55734–55784.

Zhu, Q., Luo, J., 2022. Generative pre-trained transformer for design concept generation: an exploration. Proceedings of the design society 2, 1825–1834.

Zuo, H., Jing, Q., Song, T., Sun, L., Childs, P., Chen, L., 2022. Wikilink: An encyclopedia-based semantic network for design creativity. Journal of intelligence 10, 103.