



PDF Download  
3706598.3713683.pdf  
06 January 2026  
Total Citations: 2  
Total Downloads: 963

Latest updates: <https://dl.acm.org/doi/10.1145/3706598.3713683>

RESEARCH-ARTICLE

## Creative Blends of Visual Concepts

ZHIDA SUN, Shenzhen University, Shenzhen, Guangdong, China

ZHENYAO ZHANG, Shenzhen University, Shenzhen, Guangdong, China

YUE ZHANG, Shenzhen University, Shenzhen, Guangdong, China

MIN LU, Shenzhen University, Shenzhen, Guangdong, China

DANI LISCHINSKI, Hebrew University of Jerusalem, Jerusalem, Israel

DANIEL COHEN-OR, Shenzhen University, Shenzhen, Guangdong, China

[View all](#)

**Open Access Support** provided by:

**Hebrew University of Jerusalem**

**Shenzhen University**

**Published:** 26 April 2025

[Citation in BibTeX format](#)

CHI 2025: CHI Conference on Human Factors in Computing Systems  
April 26 - May 1, 2025  
Yokohama, Japan

**Conference Sponsors:**  
SIGCHI

# Creative Blends of Visual Concepts

Zhida Sun  
CSSE, Shenzhen University  
Shenzhen, China  
zhida.sun@szu.edu.cn

Min Lu  
Shenzhen University  
Shenzhen, China  
lumin.vis@gmail.com

Zhenyao Zhang  
CSSE, Shenzhen University  
Shenzhen, China  
a3060458966@gmail.com

Dani Lischinski  
Hebrew University of Jerusalem  
Jerusalem, Israel  
danix@cs.huji.ac.il

Yue Zhang  
Shenzhen University  
Shenzhen, China  
zhangyue2020@email.szu.edu.cn

Daniel Cohen-Or  
CSSE, Shenzhen University  
Shenzhen, China  
cohenor@gmail.com

Hui Huang\*  
CSSE, Shenzhen University  
Shenzhen, China  
hhzhiyan@gmail.com



**Figure 1: The Creative Blends system leverages textual user input to generate visual blends, offering a platform for creative exploration. The system broadens the spectrum of design possibilities by mapping abstract concepts – highlighted in blue and orange colors – onto tangible physical objects through constructed metaphorical associations. These mapped objects are then seamlessly integrated based on shared attributes, resulting in conceptually meaningful and visually cohesive blends.**

## Abstract

Visual blends combine elements from two distinct visual concepts into a single, integrated image, with the goal of conveying ideas through imaginative and often thought-provoking visuals. Communicating abstract concepts through visual blends poses a series of conceptual and technical challenges. To address these challenges, we introduce *Creative Blends*, an AI-assisted design system that leverages metaphors to visually symbolize abstract concepts by blending disparate objects. Our method harnesses commonsense knowledge bases and large language models to align designers’

conceptual intent with expressive concrete objects. Additionally, we employ generative text-to-image techniques to blend visual elements through their overlapping attributes. A user study (N=24) demonstrated that our approach reduces participants’ cognitive load, fosters creativity, and enhances the metaphorical richness of visual blend ideation. We explore the potential of our method to expand visual blends to include multiple object blending and discuss the insights gained from designing with generative AI.

## CCS Concepts

• **Human-centered computing** → **User interface management systems**; *User studies*; Visualization toolkits.

## Keywords

Visual Blends, Metaphor, Text-to-Image Generation, Creativity

## ACM Reference Format:

Zhida Sun, Zhenyao Zhang, Yue Zhang, Min Lu, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. 2025. Creative Blends of Visual Concepts. In *CHI Conference on Human Factors in Computing Systems (CHI '25)*, April 26–May 01, 2025, Yokohama, Japan. ACM, New York, NY, USA, 17 pages. <https://doi.org/10.1145/3706598.3713683>

\*Corresponding author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).  
*CHI '25, Yokohama, Japan*

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
ACM ISBN 979-8-4007-1394-1/25/04  
<https://doi.org/10.1145/3706598.3713683>

## 1 Introduction

Visual blending is a powerful graphic design and communication technique, offering a creative means to convey novel and groundbreaking ideas [11, 14]. Visual blends strategically combine two distinct objects into a unified composition, achieving a harmonious effect through unique and compelling designs [10, 11, 13]. For instance, blending a dumbbell with an orange might symbolize the relationship between fitness and nutrition, highlighting how physical exercise complements proper nutritional intake in maintaining good health (refer to Figure 1 on the left). Such representations are ubiquitous in our daily lives, communicating symbolic messages from simple graphic design to intricate visual arts [2, 41]. The widespread use of visual blends stems from their capacity to engage audiences and effectively communicate complex or abstract concepts in a memorable and impactful manner.

While visual blends excel at grabbing audience attention, creating good ones presents a non-trivial design challenge. Key design considerations include selecting appropriate objects, determining suitable blending attributes, and ensuring harmonious integration of blended elements. A major conceptual challenge lies in identifying symbolic objects that effectively convey the intended message while preserving their individual recognizability [11]. From a technical standpoint, creating visually appealing and coherent blends requires meticulous attention to blending techniques and adherence to established design principles, making the process both skill-intensive and time-consuming. To convey abstract concepts or experiences, visual blends often employ metaphors [1, 11, 22], necessitating the selection of suitable source-domain objects that are both semantically relevant and visually compatible.

In the realm of Artificial Intelligence Generated Content (AIGC), the generation of images is a rapidly evolving field. Existing text-to-image (T2I) generation techniques, while capable of producing images from textual input, also provide an alternative tool for accelerating and simplifying the process of creating visual blends. However, current T2I models encounter challenges when handling abstract descriptions, as they often lead to images with distorted or nonsensical textual elements [28]. This limitation highlights the need for a deeper understanding of human cognition in transforming abstract content into visual elements in images. One intriguing aspect of addressing this problem involves incorporating Conceptual Metaphor Theory (CMT) into the image-generation process [1, 27]. By grounding abstract concepts or experiences in concrete objects and their interrelations, CMT posits that metaphor, beyond being a rhetorical device, offers a way to transform disparate concepts into more relatable and comprehensible information for users. However, current research has not yet fully explored how to represent abstract concepts in image generation, particularly when employing metaphors to combine multiple objects and their relationships.

Our research pioneers the intersection of visual metaphors and image generation, offering a unique lens for understanding the cross-modal relationships underpinning creative ideation in visual blends. We explore the potential of AIGC techniques for producing visually blended content and develop a system for semantically comparing the resulting design options. Our key innovation integrates metaphors with commonsense reasoning at both object

and attribute levels, enabling the creation of visually diverse and semantically rich blends that resonate with human metaphorical cognition. By leveraging diverse visual-textual correspondences, we aim to unlock new creative possibilities and generate blended visuals that are not confined by predefined visual paradigms (e.g., specific shapes and styles) or literal representations. Inspired by Lakoff and Johnson’s foundational work on metaphors [27], we consider language a proxy for connecting abstract ideas with visual elements, establishing a robust foundation for conceptual understanding and creative expression [45].

In line with our research, we introduce Creative Blends, an AI-powered system that assists users in generating visual blend ideas by incorporating metaphors derived from user input. Informed by interviews with eight design practitioners, we identified their needs and obstacles when creating visual blends with concrete objects, especially in drafting design options for abstract concepts. Here, “concepts” are keywords extracted from user-provided expressions, while “objects” are tangible visual elements representing these ideas. Our method leverages metaphors and image generation techniques, complemented by large language models (LLMs) and commonsense knowledge bases, to enrich the diversity and metaphoricality of design outcomes. A within-subject study with 24 participants compared Creative Blends to a baseline that combined ChatGPT (built with GPT-3.5 integrated with DALL·E 3) and Google Search. Results showed that Creative Blends significantly enhances creativity, metaphoricality, and user experience in generating visually compelling ideas for abstract concepts. Participants appreciated its ability to generate diverse outputs and inspire innovative ideas through metaphorical blends. This research underscores the value of specialized AI tools in unlocking the creative potential of generative models, bridging gaps in user understanding of AI capabilities, and expanding the possibilities of metaphor-informed visual design. The main contributions of this work are three-fold:

- We propose a metaphor-inspired approach to visual blending that identifies conceptually relevant objects through commonsense reasoning and combines them based on attribute similarity;
- We introduce Creative Blends, an AI-assisted creativity support system that transforms user-provided textual input into visually blended concepts, facilitating the exploration of diverse visual-textual correspondences;
- Through a comprehensive evaluation, we provide insights into designing abstract concepts using metaphors within the context of AIGC, showcasing the system’s potential to advance creative workflows and expand the boundaries of visual ideation.

## 2 Related Work

This section reviews prior work on metaphorical visual design, examines AI-driven image generation for creativity support, and evaluates methods for representing abstract concepts visually.

### 2.1 Metaphors in Visual Design

Metaphors are powerful tools in visual design that engage audiences by suggesting meaning rather than explicitly stating it. By linking familiar concepts (source domain) to visual elements (target domain), they promote intuitive understanding and foster deeper

audience engagement [7]. Moreover, by leaving room for interpretation, metaphors encourage active participation, prompting users to uncover nuanced connections within the visual elements [16]. This characteristic of metaphors is utilized in visual design to facilitate data comprehension and emotional connection [15]. For example, Sun et al. [44] used a postcard with a landscape painting to represent users' physical and mental health metaphorically. *Metamorphus* [47] used visual metaphors to facilitate the exploration and reflection of users' dream experiences in a meaningful way. These representations enhance the intuitive perception and interpretation of information, while metaphors serve as visual cues to facilitate understanding of underlying meaning.

Beyond their ability to engage and inform, metaphors also excel at creative communication [18], a strength evidenced by their extensive use in advertising and graphic design. Prior research has examined the structural topology of metaphors in visual representation [36], identifying three primary categories: juxtaposition, fusion, and replacement structures. In the context of fusion-based visual metaphors, Chilton et al. pioneered *VisiBlends* [11] to assist novices in creating initial blend prototypes by overlaying two objects with the same shape. They later introduced *VisiFit* [10], a tool that aids novice users in collaboratively generating visual blends through brainstorming, synthesizing, and iterating. Other research efforts have also produced creative support tools, such as *MetaMap* [25], to assist users in incorporating metaphors into their visual designs. While existing research has advanced the creation of visual metaphors, it often overlooks the diverse interpretations that specific imagery can evoke. This narrow focus limits the expressive potential of visual metaphors and constrains the creation of innovative representations of meaning. Our research aims to expand design possibilities by leveraging visual-textual correspondence to explore diverse interpretations of semantic concepts, enabling designers to explore a wider range of metaphorical visual designs.

## 2.2 Support Image Creation with Generative AI

As an emerging field, research is increasingly exploring the role of generative AI in creative image design processes. Prior work has examined the potential of generative models to enhance creativity [19, 24], laying the groundwork for our exploration of visual blend ideation. Several works focus on *divergent challenges*, aiming to expand creative possibilities by enabling users to explore diverse outputs. Tools such as *CreativeConnect* [12] promote divergent ideation by recombining visual references. Other works like *GANCollage* [48] use mood boards to encourage broad exploration of visual styles, while *DesignPrompt* [35] integrates multi-modal inputs, including text, color, and images, to expand the scope of design possibilities. This supports our goal of leveraging multimodal information to enhance idea generation, enabling greater creative expression and broader exploration of design concepts.

Another set of works addresses *convergent challenges*, catering to the need to refine and focus creative outputs by aligning generated content with user-defined goals. For example, *PromptCharm* [52] and *RePrompt* [51] primarily use text-based inputs, automating prompt refinement and providing real-time model explanations to fine-tune and align generative outputs with user intentions. *GenQuery* [42] provides iterative refinement workflows that enable users to adapt generated content through query concretization and

image modification progressively. These methods prioritize user control, addressing the challenge of balancing generative outputs with refinement to create tailored results. Despite advancements in generative models for image creation, their application in multi-object blending for design exploration is limited. Our research focuses on the ideation stage, aiming to stimulate divergent thinking through the use of image-text embeddings and commonsense knowledge, enabling an iterative exploration of a diverse range of visual possibilities driven by user input.

## 2.3 Visualizing Abstract Concepts

Abstract concepts refer to ideas or thoughts devoid of physical form or concrete qualities. They cannot be directly sensed and often represent intangible qualities, relationships, or processes [3]. Visualizing abstract concepts presents challenges due to their intangible and often subjective nature. Recent advancements leverage multimodal systems, metaphorical reasoning, and creative blending techniques to make abstract concepts more interpretable, relatable, and engaging. Prior work, such as the framework developed by Liao et al. [28], has demonstrated the potential of leveraging LLMs and T2I models to generate visual representations of abstract concepts like morality, fairness, and priority. Similarly, Liu et al. introduced *Opal* [30], a multimodal framework that deals with abstract words with recognizable subjects like symbols, showing how integrating text and visual elements fosters the understanding of abstract narratives. These works leverage T2I generation as a powerful tool for creating visual interpretations of abstract concepts.

Another body of research employs metaphors [8] to bridge the gap between abstract concepts and visual representations. For instance, *Vismantic* [54] emphasizes the use of semantic structures to guide the generation of meaningful blends, demonstrating how metaphorical visuals rooted in commonsense knowledge can enhance the interpretation of abstract ideas. Cunha et al. [13] demonstrate how blending visual elements could produce creative representations of abstract ideas, such as combining facial expressions and symbols to create new emojis that convey complex meanings. In contrast to prior approaches that focus on specific forms of blending or concrete representations, our method emphasizes expanding the exploration space for abstract concepts. This allows for flexible and diverse visual outputs, enriching the ideation process for designers and creators.

Visual blending techniques [10, 11, 21] have also been extensively explored to craft visual representations of abstract ideas. Ge and Parikh [21] advanced this field by integrating vision and language models to synthesize blended visuals, enhancing the conceptual alignment between input ideas and their outputs. *PopBlends* [50] leverages LLMs to generate conceptual blends by combining semantic elements from distinct domains. Building on existing research, our approach seeks to broaden the scope of image generation by exploring the integration of diverse objects and creating complex blending relationships between them. Utilizing a commonsense knowledge base, we systematically analyze the potential relationships within each combination, ensuring that the representation of these concepts is grounded in and aligns with human metaphorical reasoning. This also enables a more diverse exploration space and a more meaningful, relatable visual representation for abstract concept ideation.

### 3 Formative Study

To understand how designers create visual blends and uncover potential challenges, we conducted in-depth interviews with eight practitioners, four from advertising and four from graphic design. The participants included both amateur and professional designers, distinguished by whether they were engaged in full-time design-related work. We employed a semi-structured interview format to engage in open-ended discussions about a specific design task. Additionally, we used the think-aloud protocol, encouraging participants to verbalize their thoughts, concerns, and plans in detail throughout the design process to gain deeper insights.

#### 3.1 Interview Setup and Process

We recruited eight participants (5 female, 3 male) for the study, with ages ranging from 21 to 31 years ( $Mean = 24, SD = 4.04$ ). All participants had prior experience incorporating visual metaphors into their design work and held formal education in visual design. Three participants were full-time design professionals, while the remaining five were design students who integrated visual metaphors into their personal projects. During the study, participants first shared their previous experiences designing visual metaphors and were tasked with creating visual blends focused on a specific topic (i.e., “*Exercise fuels your body like vitamins*”, a theme chosen for its relatability and clarity). We emphasized the design process over the final product, requesting participants to detail the evolution of their ideas. Interview questions covered the major design activities, exploring how participants conceptualized their ideas, identified sources of inspiration, and transformed concepts into prototypes. Toward the end of the interviews, participants were asked to reflect on the most challenging aspects of the design process and to articulate their specific needs for AI tools that could streamline their workflow and enhance their creative output in the context of visual blend design.

#### 3.2 Findings

The feedback from participants revealed that they followed three main iterative processes when designing visual blends: ideation, gathering materials, and implementation. Among them, the ideation stage was cited by most participants (6 out of 8) as the most challenging and time-consuming phase. One participant (E5, Female, 21) emphasized the time-consuming nature of ideation, stating, “...*ideation is a major bottleneck. I need to brainstorm broadly to find the best solution, but this process is limited by my personal knowledge and requires constant iteration*”. According to her, the primary difficulties stemmed from selecting appropriate visual elements, identifying their commonalities, and determining how to distribute and arrange them effectively. To better understand the ideation process of creating visual blends and address the challenges and needs voiced by participants, we further subdivided ideation into three sub-steps: idea generation, concept visualization, and visual integration.

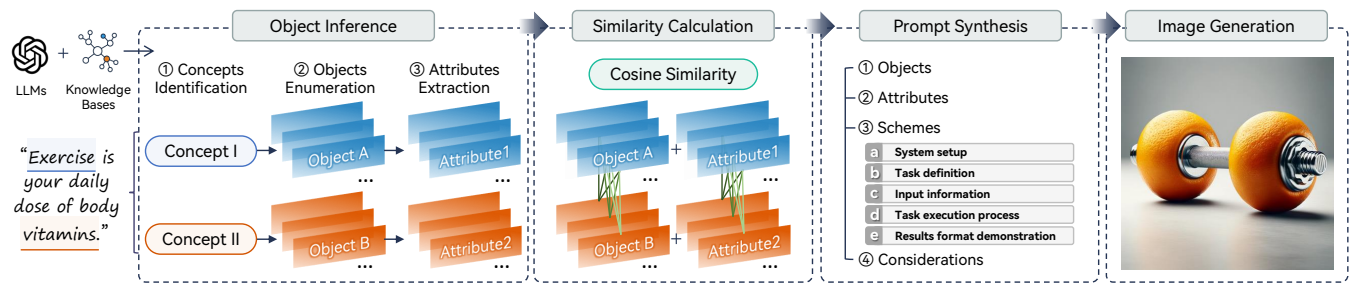
**3.2.1 Idea Generation.** Participants primarily concentrated on identifying key terms, conceptualizing the desired visual message, and choosing figurative objects that can effectively represent the expression. For all participants, keyword extraction emerged as the

universal initial step when presented with a design topic. After that, they attempted to identify appropriate expansions or elaborations of these keywords. However, several participants, including professional designers, highlighted the challenges in generating innovative design concepts, especially when dealing with ambiguous or abstract ones. One participant (E1, Female, 31) emphasized the difficulty of initiating the creative process even for experienced professionals and reflected that “...*initial concepts immediately flashed in my mind, but they often default to clichés, lacking novelty and engagement*”. Another participant (E4, Female, 28) echoed this sentiment and expressed a specific need for AI assistance in overcoming this hurdle. She envisioned an AI tool capable of generating a wide range of options as a starting point for further refinement.

**3.2.2 Concept Visualization.** Participants reported significant challenges in transforming abstract concepts into appropriate concepts related to visual representation. Identifying concrete representations for these abstract ideas was particularly time-consuming and labor-intensive, as one participant (E3, Female, 27) highlighted. This challenge intensified when design topics involved multiple abstract concepts, requiring considerable effort to visually translate and integrate them cohesively. To address this, participants expressed the need for tools or methods that could systematically examine the input abstract concepts from diverse perspectives and viewpoints. Our observations revealed that participants assessed how connected the selected concepts were when designing. Therefore, we prioritize the development of features that facilitate semantic relevance analysis for designers. This would enable a more comprehensive exploration and understanding of the concepts, better equipping the designer to create meaningful visual representations that capture the essence of the original abstract expressions.

Furthermore, participants universally recognized the importance of this step in bringing their ideas to fruition. At this step, designers are required to move beyond textual descriptions and manifest concepts visually. Two participants (E2, Female, 21; E6, Male, 21) mentioned that they would boost their creativity by taking reference for existing design solutions, as E6 reflected “*benchmarking against existing work sparks innovation and allows for iterative refinement*”. The visual output of a designer’s ideas relies on pre-defined concepts and is also a means of optimizing their design. However, designers encountered instances where suitable presentation materials for their envisioned design were unavailable, thereby restricting their creative potential. To overcome this obstacle, participants expressed a need for rapid access to a diverse range of reference images.

**3.2.3 Visual Integration.** This step involves the merging of two objects to create a cohesive and creative design. Participants identified the primary challenge, which is determining the optimal points of connection between the two objects. One participant (E8, Male, 22) commented, “*Joining two things aesthetically isn’t random; it needs a thoughtful and organized process*”. The complexity arises from the need to effectively represent individual concepts visually while maintaining a cohesive overall visual composition. By observing the general approaches designers use to merge two objects, we found that ensuring a seamless blend requires careful consideration of factors such as composition, color, perspective, texture, and scale. The importance of this step in assessing a designer’s overall skill



**Figure 2: Creative Blends operates through a multi-stage pipeline. The initial stage involves concept inference to identify relevant objects and their attributes. Subsequently, a similarity-based selection process empowers users to choose suitable object and attribute combinations. The system then explores potential blending schemes and synthesizes corresponding prompts for the T2I model, culminating in iterative image generation based on the selected prompts to support the ideation process.**

set is particularly emphasized in terms of combining disparate elements. Moreover, the process of creating visually appealing blends often involves an iterative approach of trial and error, exploring and evaluating various options throughout the creative process. Participants suggested expanding the design solution space through the divergent exploration of conceptual attributes followed by convergent selection based on attribute similarity to reduce cognitive load during the design process.

In summary, the creation of visual blends extends beyond the mere combination of two images. It requires creativity and a solid understanding of related concepts and their interconnected attributes to achieve effective and aesthetically pleasing visual communication. Based on the findings from the formative study, we summarized four design requirements as follows:

- R1** Assist with innovative idea generation while possessing clear connections to the message;
- R2** Facilitate semantic relevance analysis and visual integration based on attribute similarity;
- R3** Enable rapid access to diverse reference materials or design examples;
- R4** Allow users to try multiple ideas and compare them.

## 4 The Creative Blends System

This section introduces Creative Blends, an ideation support system designed to enhance designers' creativity by identifying and combining relevant objects and attributes for visual blending. Grounded in insights from prior research and the formative study, Creative Blends follows a multi-stage pipeline (Figure 2). The system begins by semantically analyzing user input, using LLMs and external knowledge bases to map abstract concepts to concrete objects and attributes. To aid in selecting complementary elements, it provides visualized similarity and sentiment scores, facilitating informed decision-making during the ideation process. By leveraging LLMs, Creative Blends generates detailed descriptions of potential visual blends based on user-selected inputs, blending schemes, and considerations. These descriptions are crafted into complete solutions, comprising prompts for the DALL-E 3 model, to enable rapid prototyping of diverse visual blend concepts. To support iterative creativity, users can save and revisit their outputs for refinement and

comparison. The following subsections detail the design, technical implementation, and user experience of Creative Blends.

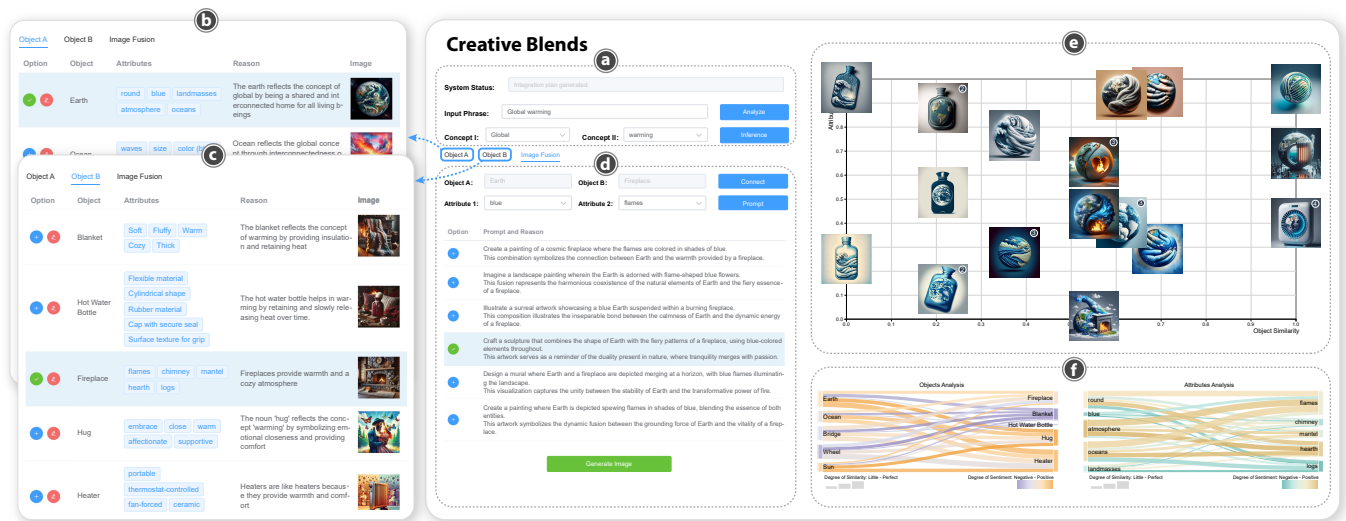
### 4.1 Design Goals

Our formative study revealed key design requirements for visual blends, emphasizing the recurring challenges designers encounter in transforming abstract concepts into visual representations through the use of metaphors. In response, we develop the Creative Blends system to support the design ideation of visual blends for a broad spectrum of abstract expressions. The system empowers designers to enhance their creativity when searching concrete imagery and enables the design to convey the underlying meaning of the expression. In light of the design requirements, the design goals of Creative Blends are as follows:

- G1 Inferring potential objects with metaphors.** Blended imagery necessitates associated objects to symbolize abstract concepts (**R1**), with the shared attributes of these objects providing a common ground for guiding the direction of the blending process (**R1**, **R2**).
- G2 Exploring similarity-based exemplars within diverse options.** The visual harmony of blended images is ensured by maintaining similarities between the objects and their attributes (**R2**). The system should support the exploration of diverse object combinations and their shared attributes to inspire creative representation (**R4**).
- G3 Offering sample designs as inspiration for creative ideation.** The system should provide sample designs based on designer-selected objects and their identified commonalities, offering valuable and timely inspiration for designers (**R3**).
- G4 Iterating the exploration on potential design choices.** The system should enable designers to easily track, compare, and iterate on blended results, encouraging creative ideation with a wide range of design possibilities (**R4**).

### 4.2 System Design

The Creative Blends system incorporates visualizations to assist users in exploring the generated visual blends. Figure 3 illustrates the user interface of Creative Blends, which processes the user input in the following manner. The system initially decomposes



**Figure 3: The user interface of Creative Blends showcases an example of generated results for “global warming”. The interface consists of four distinct modules: the expression input module (a), the prompt exploration module (b, c, d), the visual blend exploration module (e), and the similarity visualization module (f).**

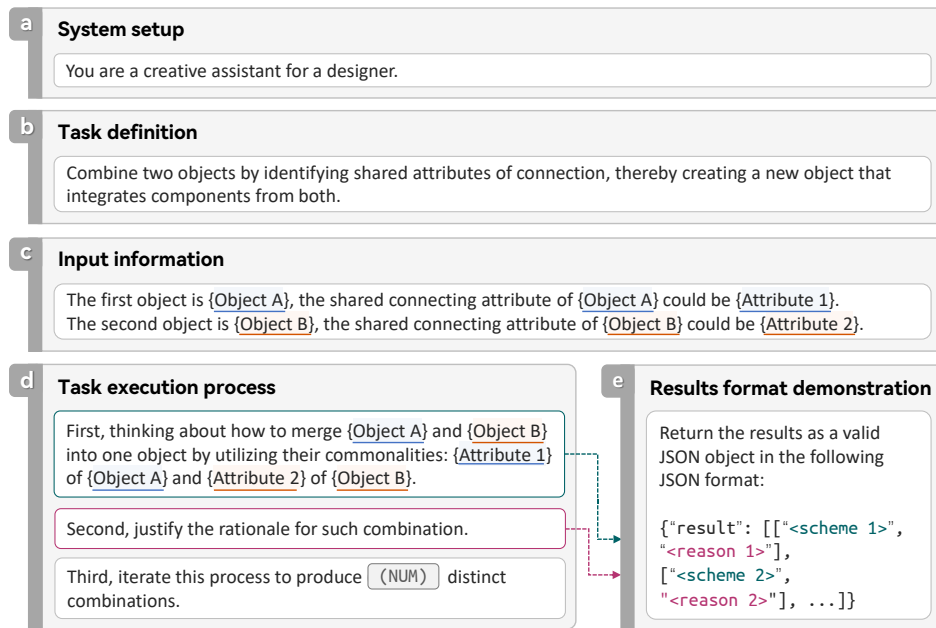
the provided expression into its constituent parts of speech. Users are then prompted to select the concepts they are interested in for further exploration. Based on the selected results, the system identifies related objects that can be metaphorically linked to the chosen concepts and extracts their physical attributes (G1). To facilitate the exploration of object and attribute combinations, Creative Blends employs Sankey diagrams to visualize semantic and similarity relationships. The diagram’s node-link representation, coupled with color-coded edges, enhances data clarity and readability, empowering users to explore diverse combinations and foster creative ideation (G2). Additionally, the system generates tailored prompts based on the user-selected objects and attributes, providing guidance for the blending process. Other relevant factors, such as the intended expression, chosen metaphors, and applicable design restrictions, are also incorporated into these prompts, which serve as input for T2I models (G3). Finally, the system generates and displays the blended results on a 2D canvas, enabling users to explore, compare, and identify the preferred ideation outcomes (G4).

### 4.3 Implementation

**4.3.1 Identifying Metaphorical Objects and Their Attributes.** To achieve our design goals, we initially tokenize the user input into adjectives, nouns, and verbs, allowing users to select the appropriate concepts from them (Figure 3a). The entire user input will later provide the model with a comprehensive context, allowing it to understand the intended concept better and identify any underlying metaphors. Next, we integrate ConceptNet [43], an external knowledge base, with GPT-3.5-turbo [5] to assist users in discovering objects connected to the given concepts. Leveraging CMT, we embed the metaphorical expression template, “{Concept} is like [a/an] {Object}”, into the prompts to identify related objects (G1). Specifically, ConceptNet is leveraged to extract the top

50 semantically related objects for the target concept, which are then incorporated into the prompt to guide GPT in selecting the most appropriate ones. For each object, ConceptNet is further utilized to identify associated attributes, which are integrated into the GPT prompts as contextual references, enhancing its capacity to suggest relevant attributes. During each query iteration, the system presents five objects to the user, enabling iterative refinement and updates until a suitable selection is achieved. Subsequently, users can review the rationale behind the metaphorical connections between each object and its associated concept, along with the top five attributes identified for each concept-object pair (Figure 3b&c). As shown in Figure 1 (left), if the user selects “vitamins” as a concept, the system suggests related objects like “orange”, “medicine”, and “egg”. Upon choosing “orange” as the object, GPT suggests that the metaphorical connection between “vitamins” and “orange” arises from the perspective that “it contains vitamin C, which is essential for health and well-being”. Additionally, the system identifies related attributes of “orange”, such as “round” and “orange color”.

**4.3.2 Calculation of Similarity and Sentiment Scores.** To facilitate the exploration of various blending options, we incorporate similarity and sentiment scores. These scores help users identify and compare similarities among different objects or attributes and understand the overall harmony of blended results. We leverage a pre-trained CLIP model [38], which encodes both image and text data into a shared embedding space, thereby capturing the semantic relationship between textual and visual modalities. By computing cosine similarity in the CLIP text embedding space [39], we assess the similarity among objects and evaluate their likeness. The similarities between attributes can also help designers identify commonalities among objects, which can serve as anchors during the blending process (G2). The Sankey diagrams represent the similarity scores between objects or attributes using the width of their



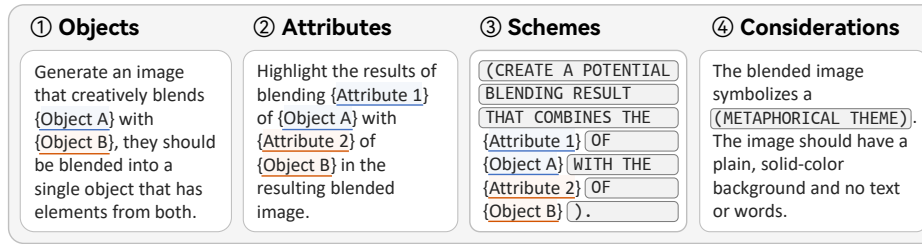
**Figure 4: The scheme generation prompt is structured into five key modules: (a) system setup, (b) task definition, (c) user input, (d) task execution process, and (e) results format demonstration. The (d) task execution process module outlines the methods and rationales for considering potential blending schemes. The results, processed by the (e) results format demonstration module, are returned in a standardized JSON format, ensuring compatibility with downstream processes.**

connecting links (Figure 3f), making it easier for users to understand the relationships between different options and aiding them in observing and comparing various generation outcomes (G4). To preserve the relationships among the original data values and improve the visual clarity of the similarity matrix, we employ the Min-Max normalization [6] to standardize the similarity scores.

We use the DistilBERT model [40] to perform sentiment analysis on text descriptions of objects or attributes, thereby assisting users in making informed selections and reducing the likelihood of misinterpretations due to contextual variations or other factors. Sentiment scores are calculated using the confidence level  $C$  of the sentiment label, where “positive” labels are equal to  $C$  while “negative” labels are calculated as  $(1 - C)$ . We employ color variations to distinguish between objects and attributes while ensuring that the color temperature gradient correlates with semantic sentiment changes. The Sankey diagrams use purple and orange to represent negative and positive sentiments for objects, and green and gold for attributes, respectively. To determine the overall sentiment between two objects or attributes, we average their individual sentiment scores. Subsequently, we apply quantile normalization to the overall sentiment scores to ensure that the data distributions are similar across different samples. This also makes the color distribution in the Sankey diagrams more balanced.

**4.3.3 Generate Blending Prompts and Images.** Following the user’s selection of objects and attributes, Creative Blends generates blended images by first establishing blending schemes (Figure 3d), and then converting the chosen scheme into a final prompt for image generation (G3). To enhance user control and creative freedom, we

adhered to prompt engineering best practices [31, 37] for blend scheme construction by outlining the GPT’s role, the task, the user input, and a step-by-step process to guide the model (Figure 4). Subsequently, the user’s selected scheme, objects, attributes, and design considerations (e.g., metaphorical themes and design constraints) were integrated into the final prompt for visual blend generation (Figure 5). These design considerations were refined through iterative experimentation. The complete prompt scripts used in our system are available in the supplementary material. Finally, we use DALL·E 3 to generate images based on the prompt, swiftly transforming designers’ creativity into reality. The generated visual blends are displayed in the image exploration area (Figure 3e). Users can explore these results based on object and attribute similarity (X and Y axes), uncovering new creative possibilities for visual blend creation (G4). To provide a clearer overview of the generated content and support iterative exploration, we number each image group with the total number of images produced for that prompt, displayed in the upper-right corner. Furthermore, to maximize design diversity, we avoid imposing unnecessary visual constraints on blend generation beyond the specified design considerations. This approach leads to a diverse range of design outcomes within the 2D exploration space. For prompts that may be closely related, Creative Blends implements interactive zooming to adjust the scale of the visualization space and prevent image overlap. Our system also facilitates diverse and iterative creative exploration by allowing users to alter specific objects when they conceive improved or alternative ideas (G4). Users can directly replace the objects in the system, prompting it to regenerate new images while discarding all prior visual information.



**Figure 5: The final prompt is composed of four distinct modules: (1) objects, (2) attributes, (3) schemes, and (4) considerations. The schemes and metaphorical themes (marked in capital letters with a grey background), essential elements of the prompt, are dynamically generated by the GPT in response to the scheme and metaphor generation prompts.**

#### 4.4 User Scenario

This section demonstrates how Creative Blends assists designers in creating visual blends. Alice, a graphic designer, wants to create an image that describes the concept of “global warming” to depict the idea of environmental protection. She starts by entering this phrase into the system (Figure 3a). Creative Blends parses the input and identifies keywords like “global” and “warming”. Alice selects these keywords to discover specific objects that can metaphorically represent them. The Creative Blends system presents two lists of objects, along with their associated attributes (Figure 3b&c). Two Sankey diagrams visually illustrate the connections between these objects and attributes (Figure 3f). Alice initially examines the object names, their relevant attributes, and how they align with her concept (Figure 3b&c). For unfamiliar objects, she can click the “preview” button to view generated images for more information. She then examines the “Objects Analysis” Sankey diagram, which aids in object selection by displaying similarity and sentiment scores. As Alice explores the Sankey diagram, clicking a link automatically selects the corresponding pair of objects in the lists (Figure 3b&c). She finally picks “earth” and “fireplace” as positively oriented objects with a higher degree of similarity.

Alice then delves into the system’s image-blending capabilities, beginning with an examination of the “Attributes Analysis” Sankey diagram for guidance. This diagram visually represents the similarity and sentiment scores among the objects’ attributes. By referencing similarity, Alice can assess the coherence of the current blending results and plan her next steps. She experiments with various attribute combinations to adjust how the objects blend within the prompt. Ultimately, she confirms the attributes “round” for “earth” and “flames” for “fireplace”. Once satisfied with her combinations, she generates the corresponding scheme prompts along with the previously selected objects (Figure 3d). The final prompt then creates a new blended image in the display area, which she can enlarge and examine (Figure 3e). When multiple images are created at the same location, the system displays the total image count in the upper right corner. After several attempts, Alice realizes she has yet to explore certain design options for visual blending. Inspired by her exploration, she decides to utilize the editing feature to modify an object from “fireplace” to “ice cream” (Figure 3c). With further experimentation and exposure to additional sample designs of visual blends with different combinations, Alice gains new ideas that she can incorporate into her subsequent creations.

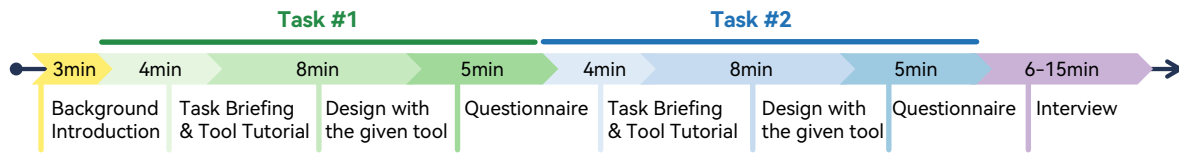
#### 5 Hypotheses

Previous studies indicate that creating visual blends is a challenging task [10, 11]. However, a broader range of related examples can enhance the ideation process [17]. Additionally, reducing the cognitive load unrelated to creative tasks can boost creativity [32]. Given the specific application of our system and the creative practices of designers, we propose the following hypotheses relative to the baseline (described in section 6):

- H1** Compared to the baseline, Creative Blends improves the overall usability by streamlining the design ideation workflow.
- H2** Compared to the baseline, Creative Blends is capable of decreasing users’ mental demand (**H2a**), physical exertion (**H2b**), and time pressure (**H2c**) during visual blend design ideation, while simultaneously improving their satisfaction with the design ideation outcomes (**H2d**), reducing the required effort (**H2e**), and mitigating frustration during task execution (**H2f**).
- H3** Compared to the baseline, Creative Blends is capable of generating a greater quantity (**H3a**) and a wider variety of design ideation outcomes (**H3b**) in the same amount of time, leading to improved overall designer satisfaction (**H3c**).
- H4** Compared to the baseline, Creative Blends enables users to explore a broader range of creative ideas more easily (**H4a**), facilitates collaboration between designers and AI (**H4b**), and enhances the enjoyment of the design process (**H4c**). Additionally, users are likely to perceive the generated ideation results as more worthwhile (**H4d**), and the system’s user interface enables them to focus on the task itself (**H4e**). Furthermore, users can exhibit increased expressiveness during their design ideation activities (**H4f**).
- H5** Compared to the baseline, Creative Blends offers a more metaphorical approach to supporting user ideation activities.

#### 6 Evaluation

To test our hypotheses, we conducted a controlled experiment with 24 participants using a within-subject design to minimize individual differences. To ensure a fair comparison and avoid the influence of varying image generation speed and style from different models, we opted for a more rigorous approach by using ChatGPT (GPT-3.5 with DALL-E 3) and Google Search as the baseline. We focused on their core features, such as information search, image



**Figure 6: The user study procedure. It involved completing two tasks with both the Creative Blends and the baseline. To maintain fairness, the order of the systems and design tasks was counterbalanced.**

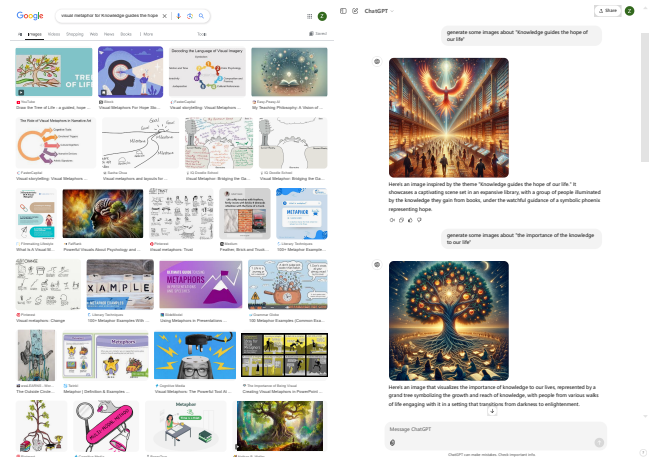
generation, and browsing history. Participants undertook design ideation tasks across the Creative Blends and the baseline. Clear instructions were provided at the outset of each task to enable participants to explore and interact with both conditions freely. After completing each task, we administered questionnaires to gather participants’ evaluations of their experiences with the respective system conditions. Our questionnaire design was informed by relevant literature [4, 9, 23] and guided by our research hypotheses, which shaped the evaluation around the following dimensions: we employed the System Usability Scale (SUS) and NASA-Task Load Index (NASA-TLX) to gauge the system’s usability and the cognitive load imposed on users. Additionally, we incorporated the Creative Support Index (CSI) and items focused on outcome satisfaction and metaphoricity to measure the extent to which the system fostered creative ideation, user satisfaction, and the integration of metaphor within the design process. Figure 6 depicts the complete evaluation process for each participant.

## 6.1 Participants

We enlisted a total of 24 participants (12 female, 12 male) aged 20 to 31 ( $Mean = 24.08, SD = 3.31$ ) through online advertising and word-of-mouth. The participants were local university students and faculty members, with equal representation from both computer science (6 female, 6 male) and design (6 female, 6 male) backgrounds. Our selection criteria included a demonstrated interest in visual design and some prior experience in design activities [25]. Additionally, all participants possessed normal color vision and had experience with conversational AI and T2I models.

## 6.2 Tasks and Procedure

During the experiment, each participant was tasked with producing visual blending ideas for the two abstract expressions: “Smoking is like a warm welcome to death” (T1) and “Knowledge guides the hope of our life” (T2). Based on input from professional designers, our criteria for selecting these two topics included the presence of at least two concepts, their abstract nature that allows visualization with specific objects and attributes, the potential for considerable exploration, and the inclusion of associated imagery that is either positive or negative. Participants were free to explore their creativity by creating visual blends related to the given topics. To maintain experimental balance, we employed a Latin Square design, yielding four possible sequences: (a) T1 (Baseline) - T2 (Creative Blends), (b) T2 (Baseline) - T1 (Creative Blends), (c) T1 (Creative Blends) - T2 (Baseline), and (d) T2 (Creative Blends) - T1 (Baseline).



**Figure 7: The baseline interface integrated Google Search and ChatGPT for both text and visual search queries.**

Prior to the main study, we spent three minutes explaining the basic concepts of visual blends, encouraging participants to incorporate visual metaphors into their creative process. The main study had two parts, each using either the baseline or Creative Blends system. The baseline interface integrated Google Search and ChatGPT into a unified side-by-side display (Figure 7), enabling users to interact with both platforms simultaneously. Each task began with a 4-minute tutorial, offering an overview of the design process and step-by-step instructions, followed by a Q&A session for clarification. To mitigate order effects, task sequences were rotated using a Latin square design. To ensure a fair evaluation, we provide users with sample prompt words as a reference when introducing the baseline condition. After the tutorials, participants completed design tasks using the baseline or Creative Blends system within 8 minutes. Once finished, they filled out 5-minute questionnaires about their design experience. We also conducted 6- to 15-minute interviews to gain a deeper understanding of their experiences. The experiment ended after the interviews, and the total duration did not exceed 60 minutes.

## 6.3 Results Analysis

This section presents a statistical analysis of user ratings gathered during the experimental process. The findings, contextualized by user feedback, are presented in relation to each evaluation criterion.

**6.3.1 System Usability.** To assess system usability, we adopted the System Usability Scale (SUS) [4] as our questionnaire. The means

**Table 1: The statistical results of user feedback with the Creative Blends and the baseline (i.e., control group), where the p-values (-:  $p > .100$ , +:  $.050 < p < .100$ , \*:  $p < .050$ , \*\*:  $p < .010$ , \*\*\*:  $p < .001$ ) is reported.**

Category	Factor	Creative Blends		Baseline		Statistics				Hypotheses
		Mean	SD	Mean	SD	<i>t</i>	<i>Z</i>	<i>p</i>	Sig.	
<b>Usability</b> [4]	\	75.000	9.918	61.458	21.781	2.547	\	0.018	*	<b>H1 Accepted</b>
<b>Cognitive Load</b> [23]	Mental Demand	2.542	1.062	4.458	1.587	-4.861	\	<0.001	***	<b>H2a Accepted</b>
	Physical Demand	1.625	0.824	2.958	1.967	\	-2.865	0.004	**	<b>H2b Accepted</b>
	Temporal Demand	3.250	1.359	3.833	1.685	\	-1.517	0.129	-	<b>H2c Rejected</b>
	Performance	4.583	0.929	4.333	1.736	0.598	\	0.556	-	<b>H2d Rejected</b>
	Effort	2.583	1.349	3.917	1.666	-2.734	\	0.012	*	<b>H2e Accepted</b>
<b>Outcome Satisfaction</b>	Frustration	2.625	1.663	3.625	1.663	-1.931	\	0.066	+	<b>H2f Rejected</b>
	Amount	5.080	1.412	4.540	1.444	1.299	\	0.207	-	<b>H3a Rejected</b>
	Diversity	5.380	1.345	3.670	1.274	5.058	\	<0.001	***	<b>H3b Accepted</b>
<b>Creativity</b> [9]	Overall	5.080	1.283	3.960	1.367	2.609	\	0.016	*	<b>H3c Accepted</b>
	Exploration	5.708	1.160	4.083	1.886	2.985	\	0.007	**	<b>H4a Accepted</b>
	Collaboration	5.500	1.560	3.667	2.014	3.350	\	0.003	**	<b>H4b Accepted</b>
	Enjoyment	5.833	1.435	4.042	1.756	3.558	\	0.002	**	<b>H4c Accepted</b>
	Results Worth Effort	5.458	1.444	4.167	1.633	2.806	\	0.010	*	<b>H4d Accepted</b>
	Immersion	5.167	1.903	5.042	1.706	0.213	\	0.833	-	<b>H4e Rejected</b>
<b>Metaphoricity</b>	\	5.167	1.167	4.375	1.689	1.807	\	0.084	+	<b>H4f Rejected</b>
<b>Metaphoricity</b>	\	5.655	0.974	5.047	1.193	2.276	\	0.032	*	<b>H5 Accepted</b>

and standard deviations of the participants' ratings are presented in Table 1. From the participants' feedback, a Shapiro-Wilk test is conducted to examine the distribution's normality, and no evidence of non-normality is observed in the overall distribution ( $W = 0.925, p = 0.075$ ). A paired samples t-test comparing Creative Blends to the baseline shows significantly higher usability for Creative Blends ( $t(23) = -2.547, p = 0.018$ ). The effect size of Cohen's  $d$  ( $d = 0.800$ ) suggests a large effect on the improvement in Creative Blends usability. Therefore, **H1** is accepted. Two participants (P9, Male, 21; P17, Male, 21) praised Creative Blends for its simple and user-friendly interface, describing it as "refreshing". Almost all participants (22/24) found Creative Blends relatively easy to learn. As one participant (P18, Female, 23) noted, "While I needed some help at first to use Creative Blends, I was able to use it independently after practicing once." Unlike the conversational interface of the baseline, Creative Blends presents generated results via a 2D mood board-like interface (Figure 3e), providing a comprehensive overview of the output's scope. The sentiment analysis visualizations also empower users to evaluate and refine the blending process effectively.

**6.3.2 Cognitive Load.** To evaluate cognitive load, we referred to the NASA-Task Load Index (TLX) [23]. The questionnaire is tailored to encompass six factors: mental demand (the amount of mental effort required to complete a task), physical demand (the degree of physical exertion needed to complete the task), temporal demand (the time taken to complete the task), performance (the impact of task completion), effort (the amount of effort necessary to complete the task), and frustration (the dissatisfaction experienced while completing a task). Table 1 provides the mean and standard deviation of the participants' ratings. The Shapiro-Wilk test reveals no evidence of non-normality in the distribution of mental demand ( $W = 0.594, p = 0.967$ ), performance ( $W = 0.948, p = 0.246$ ), effort ( $W = 0.962, p = 0.479$ ), and frustration ( $W = 0.948, p = 0.243$ ) factors. Therefore, paired samples t-tests were employed to compare the differences in these four dimensions between the two systems,

as shown in Table 1. However, the distributions of physical demand ( $W = 0.811, p < 0.001$ ) and temporal demand ( $W = 0.870, p = 0.005$ ) factors deviate from normality. Consequently, the differences between the two systems in these dimensions were compared using the Wilcoxon signed-rank test.

The test results indicate that the cognitive load in the dimensions of mental demand ( $t(23) = 4.861, p < 0.001$ ), physical demand ( $Z = -2.865, p = 0.004$ ), and effort ( $t(23) = 2.734, p = 0.012$ ) when using Creative Blends is significantly lower than that of the baseline. The effect sizes of Cohen's  $d$  and Pearson's correlation coefficient  $r$  indicate that the reduction in mental demand ( $d = -1.419$ ) was large, physical demand ( $r = -0.414$ ) was moderate, and effort ( $d = -0.880$ ) was large. Consequently, hypotheses **H2a**, **H2b**, and **H2e** are accepted. However, no significant differences were found in the dimensions of temporal demand, performance, and frustration, leading to the rejection of hypotheses **H2c**, **H2d**, and **H2f**. The results show that there is little difference in the time spent and performance of the generated results between the Creative Blends and the baseline, as they both use the same underlying technologies. However, our approach significantly reduces the mental and physical burden on users and requires less effort from them. Four participants (P1, Female, 31; P2, Male, 23; P14, Male, 22; P23, Female, 30) praised the pipeline and interface applied in Creative Blends. They appreciated the time-saving benefits of not having to initiate the design process from scratch, and the generated results were more controllable, making their efforts more manageable. Other participants further emphasized the value of the Sankey diagrams in streamlining the selection of appropriate object and attribute pairings, highlighting their ability to provide an overview and thereby minimize the cognitive effort required for individual pair evaluations.

**6.3.3 Outcome Satisfaction.** To measure outcome satisfaction, we designed the questionnaire inspired by MetaMap [25]. The questionnaire includes three factors, i.e., satisfaction with the amount,

diversity, and overall quality of the system output. We gathered participants' ratings and conducted a Shapiro-Wilk test. The results show no evidence of non-normality in the distribution of outcome satisfaction ratings for the amount ( $W = 0.939, p = 0.154$ ), diversity ( $W = 0.931, p = 0.101$ ), and overall quality ( $W = 0.936, p = 0.129$ ). Therefore, paired samples t-tests were employed to compare the differences in outcome satisfaction between the two systems. Table 1 provides the mean and standard deviation of the relevant factors.

Overall, participants were significantly more satisfied with the outcomes from Creative Blends than with those from the baseline ( $t(23) = -2.609, p = 0.016$ ). This large effect size (Cohen's  $d = 0.845$ ) indicates a meaningful improvement in satisfaction with Creative Blends. Specifically, participants were significantly more satisfied with the diversity of outcomes generated by Creative Blends ( $t(23) = -5.058, p < 0.001; d = 1.305$ ). However, there was no significant difference between the two systems in the number of outcomes generated. Therefore, hypotheses **H3b** and **H3c** are accepted, while hypothesis **H3a** is rejected.

Participants indicated that the Creative Blends system facilitated the generation of more varied, predictable, and actionable expressions of their ideas. Nearly half of the participants (11 out of 24) mentioned that the baseline often produced outcomes that deviated from their intended concepts, while Creative Blends guided users in expanding their design choices systematically. Some participants expressed frustration with the baseline, stating, "I feel like my prompts are pretty clear, but I am struggling to get good results...It is really frustrating that these results are not even close to what I am looking for" (P8, Female, 21). In contrast, P17 (Male, 21) commented, "I find it much easier to achieve the desired effect with Creative Blends. You can start with one idea and branch off into lots of different directions". The association-based thought expansion approach for visual blend design, which links objects and their attributes, has demonstrated its effectiveness in rapidly generating diverse results.

**6.3.4 Creativity.** We borrowed the Creative Support Index (CSI) [9] to assess creativity. We modified the CSI questionnaire to include six metrics: exploration (the extent to which the system supports user exploration), collaboration (the manner in which users collaborate with AI), enjoyment (the level of enjoyment participants experienced during the activity), results worth effort (the value derived from the effort), immersion (the degree of focus within the system), and expressiveness (the expressiveness and creativity exhibited during the activity). Table 1 provides the mean and standard deviation of these six factors. The Shapiro-Wilk tests indicate that the distributions of scores for all six factors are normal. For each factor, the following values were obtained: exploration ( $W = 0.944, p = 0.205$ ), collaboration ( $W = 0.964, p = 0.524$ ), enjoyment ( $W = 0.946, p = 0.225$ ), results worth effort ( $W = 0.946, p = 0.227$ ), immersion ( $W = 0.975, p = 0.791$ ), and expressiveness ( $W = 0.953, p = 0.307$ ).

The results in Table 1 show that Creative Blends significantly outperforms the baseline in supporting creativity in exploration ( $t(23) = -2.985, p = 0.007$ ), collaboration ( $t(23) = -3.350, p = 0.003$ ), enjoyment ( $t(23) = -3.558, p = 0.002$ ), and results worth effort ( $t(23) = -2.806, p = 0.010$ ). The substantial advantages of Creative Blends are reflected in the large effect sizes, with Cohen's  $d$  values for exploration ( $d = 1.038$ ), collaboration ( $d = 1.018$ ), enjoyment ( $d = 1.117$ ), and results worth effort ( $d = 0.838$ ). Compared to

the baseline, participants found Creative Blends to be more supportive of exploring, collaborating, and enjoying the creative process. Additionally, participants perceived a significant improvement in the value of their efforts when using Creative Blends. Consequently, hypotheses **H4a**, **H4b**, **H4c**, and **H4d** are supported, whereas hypotheses **H4e** and **H4f** are refuted.

Feedback from participants suggests that our system significantly improves the exploration of design options, enhances engagement, and reduces the burden of ideation during the creative process. Five participants emphasized that Creative Blends helped them express abstract concepts in visual representation more effectively than the baseline (P3, Male, 27; P11, Male, 27; P13, Female, 25; P19, Male, 22; P22, Male, 26). As P19 (Male, 22) reflected, "...*(Creative Blends) could really help me think outside the box and come up with some new solutions that I would not have thought of on my own*". Creative Blends leverages metaphors to facilitate inter-domain concept transfer. By examining the underlying relationships between disparate concepts, the system produces a more multifaceted assortment of object combinations. Additionally, our approach fosters collaboration between users and AI. Like P1 (Female, 31) remarked, "...*basically, I tell Creative Blends what I am looking for, and it fills in all the gaps. It is like giving AI a blueprint, and then it builds the whole thing*". These findings provide empirical evidence that human-AI collaboration extends beyond simple task delegation, encompassing the ability of humans to decompose complex problems into manageable subtasks for AI processing effectively. Moreover, two participants stated that Creative Blends increased their ability and willingness to explore (P18, Female, 23; P23, Female, 30). Even after the design session ended, both expressed a desire to continue exploring.

**6.3.5 Metaphoricity.** To assess the metaphorical quality of the generated outputs, we adopted a method outlined in previous research [49]. Participants were asked to indicate the extent to which they perceived a metaphorical relationship between the target and source elements, using a 7-point Likert scale ranging from "no metaphor" to "extremely strong metaphor". The Shapiro-Wilk test confirms the normality of the distribution ( $W = 0.935, p = 0.124$ ).



**Figure 8: Eight sample outputs generated by the Creative Blends and the baseline for topics of T1 and T2.**



**Figure 9: Creative Blends generates diverse visual blends representing abstract concepts based on user-provided expressions. Each topic includes eight examples: four highlighting different levels of object similarity and four demonstrating varying attribute similarity. Similarity increases from left to right. The attributes are extended based on the objects enclosed by the double brackets. Colors within the topics serve to identify concepts and their associated objects and attributes.**

A paired samples t-test reveals a significant difference in overall similarity between Creative Blends and the baseline ( $t(23) = -2.276, p = 0.032$ ). The moderate effect size (Cohen’s  $d = 0.558$ ) indicates a meaningful improvement in the metaphorical quality of content generated by Creative Blends, leading to the acceptance of hypothesis **H5**.

Participants’ feedback highlights the significance of metaphor in bridging the gap between abstract concepts and familiar imagery. As P15 (Male, 23) observed, “... the (baseline) often turns abstract information into scenes rather than specific characters”. This shows that existing models are not yet able to associate abstract meanings with existing imagery of physical objects. In addition, six participants noted that the target and source elements within Creative Blends outcomes were more readily recognizable compared to the baseline (P5, Male, 21; P7, Female, 21; P18, Female, 23; P21, Male, 20; P22, Male, 26; P24, Female, 23). These findings suggest that the Creative Blends system preserves the original features of blended objects while aligning metaphorical relationships with human cognition within the constraints of the commonsense knowledge derived from the physical world.

## 6.4 Ideation Results

We collected the outputs produced by participants during the study. These results have been documented and are provided as part of

our supplementary materials. Figure 8 presents eight representative examples selected from four participants (i.e., P1, P5, P6, and P18), showcasing typical outcomes generated by both system conditions. Our analysis revealed that both conditions are capable of generating images that use visual metaphors to represent abstract concepts. However, Creative Blends exhibited a notable ability to generate blended results that align with the definition of visual blends (as shown in Figure 8). In contrast, the baseline often relied on background elements, scene composition, or visual juxtaposition to communicate abstract ideas, and, in some cases, simply incorporated related text directly within the images.

To demonstrate the capability of Creative Blends, we showcase generated results for four additional topics in Figure 9. These examples illustrate the versatility of Creative Blends’s output. The examples in the left part of Figure 9 show the system’s ability to explore physical objects that vary in their relevance to abstract concepts. To achieve optimal object combinations, users can adjust the blending structure and connection method according to the similarity of their attributes (Figure 9 right). Overall, Creative Blends facilitates metaphorical mapping between physical objects and their associated abstract concepts, and builds connections between these different objects, stimulating creative ideas for visual blend creation.



**Figure 10: Designers have two options when choosing visual metaphors: explicit (left) or implicit (right).**

## 7 Discussion

This section delves into user preferences for creating visual blends and the generalizability of our method. We also provide insights for designing with generative AI and highlight the study’s limitations.

### 7.1 Design Preferences of Visual Blends

During the evaluation of Creative Blends, we observed a variety of user designs reflecting their individual preferences. These preferences were apparent not only in our interviews but also in how they interacted with both systems. We can categorize these preferences into three main design options: a preference for the visual expression of metaphors, the features prioritized for blending, and the visual style of blended results.

**7.1.1 Explicit vs. Implicit Expression of Visual Metaphors.** The design content of visual blends can directly or indirectly represent the topic. For instance, in Figure 10, when creating visual blends for “global warming”, we can choose the “fireplace” from the explicit direction to establish a direct metaphorical connection with the concept of “warming”. Alternatively, we can select the “ice cream” from the implicit direction to represent this concept through its melting state. Similarly, for the task “Pepper sauce: be aware of the heat”, we can opt for the “lighter” from the explicit direction to directly represent the concept of “heat”. Or, we can choose the “fire extinguisher” from the implicit direction to represent this concept through its fire-extinguishing or heat-reducing function.

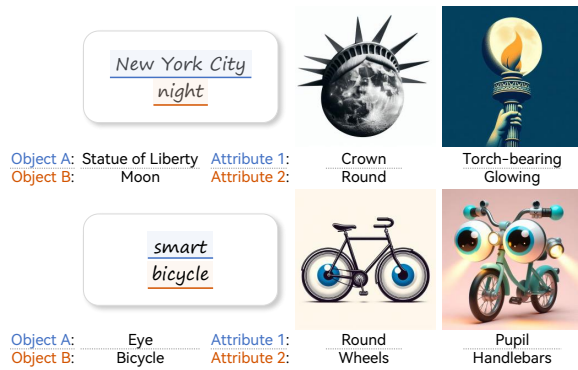
Almost half of the participants (13 out of 24) preferred implicit expression, finding it more subtle (P20, Female, 31; P24, Female, 23) and thought-provoking (P13, Female, 25). This approach allows the audience to gain a deeper understanding of the design output. These advantages contribute to the engagement (P11, Male, 27) and imagination (P9, Male, 21) of the resulting outcomes (P12, Female, 24). P9 (Male, 21) noted that implicit expressions provide richer metaphorical information. While implicit expressions offer these

benefits, explicit visual metaphors succeed in directly representing the literal information, providing a clear and straightforward understanding.

**7.1.2 Prioritizing Semantic vs. Visual Features when Blending.** When visually representing abstract concepts, a debate exists regarding the optimal object features to utilize. Participants’ preferences were evenly distributed, with eight out of 24 participants favoring semantic features, while nine favored visual features. Those who preferred semantic features believed that semantic meaning captures the contextual nuances of concepts, enabling accurate representation (P20, Female, 31) and facilitating the understanding of abstract concepts and relationships that are not immediately apparent visually (P18, Female, 23). They hypothesized that blended results could convey the topic more directly, clearly, and comprehensively (P15, Male, 23). However, participants also recognized the potential ambiguity of semantic meanings arising from polysemy (words with multiple meanings) and homonymy (different words that share the same spelling). Conversely, those who preferred visual features favored a straightforward utilization of visual elements such as color, shape, and texture when creating visual blends (P8, Female, 21; P14, Male, 22). They believed that higher visual similarity would result in more harmonious blended visual outcomes (P6, Female, 23). Nevertheless, the superficiality of visual representation can also lead to misinterpretations of meaning. We propose that when creating visual blends, designers should seek suitable objects based on semantic features and then identify the most appropriate forms of related objects for blending using visual features to achieve more effective conceptual expression.

Beyond using relevant features to facilitate harmonious blending, similarity scores within each feature also play a pivotal role in guiding designers toward diverse exploration paths. While high similarity scores are often desirable, designers may intentionally select object combinations with lower similarity scores to achieve more exaggerated and visually striking results. As illustrated by the example “exercise is your daily dose of body vitamins” in Figure 9, “orange” and “basketball” exhibit greater visual compatibility than “orange” and “badminton” and are easier to combine; however, this may lead to a less innovative visual outcome. This approach encourages designers to explore a broader range of possibilities, stimulating their creativity and accelerating the ideation process.

**7.1.3 The Integrated vs. Independent Styles of Blended Outcomes.** Upon analyzing the generated visual blends, we identified two primary blending styles: objects merging into a unified form while retaining their individual identities or remaining independent and connected by shared attributes. For instance, the visual blends of “global warming” (first row) in Figure 9 exhibit a cohesive appearance between different objects. The integrated representation creates a seamless and unified visual, which can be aesthetically pleasing and easier to interpret as a single entity. However, achieving a balanced and harmonious blend can be challenging and may necessitate a thorough analysis of the objects’ attributes. In contrast, the visual blends of “smoking is like a warm welcome to death” (last row) in Figure 9 present the distinct characteristics of each object. Such independent representation allows for more flexibility in design, as objects can be repositioned or modified independently without affecting the overall composition. Using common attributes



**Figure 11: Creative Blends draws upon topics from *VisiBlends* [11] to showcase the related physical objects and the interconnections among their attributes.**

to connect objects can highlight the relationships or interactions among them, providing additional information. However, the separated objects may not convey a sense of harmony and unity as effectively as the integrated design, and an excessive focus on individual characteristics might detract from the overall message. Overall, the visual blend’s style is largely shaped by the attributes of the combined objects. Designers have the flexibility to optimize the representation by substituting associated objects or adjusting visual attributes.

## 7.2 Generalization of Creative Blends

Creative Blends is designed and implemented to create visual blends by merging two objects representing distinct concepts. To test its generalizability, we applied the system to additional topics studied in prior research and expanded our experiments to evaluate its ability to combine multiple objects.

**7.2.1 Create Visual Blends with Previous Topics.** To demonstrate Creative Blends’s performance in generating visual blends, we employed the same topics used in *VisiBlends* [11]. Unlike the manual brainstorming and image selection processes in *VisiBlends*, our system automatically identifies metaphorically related objects and offers a variety of blending options based on shared attributes. This facilitates creative ideation and reduces user effort. Figure 11 demonstrates two topics from *VisiBlends*. Beyond the high-quality visual results achieved through advanced image generation techniques, our approach provides a broader range of blending possibilities. By considering attribute combinations rather than solely relying on object shapes, we enable more creative blending directions. Additionally, our method focuses on the essential aspects of physical objects that connect to abstract concepts, leading to natural and visually appealing blended outcomes.

**7.2.2 Increase the Number of Blended Objects.** When blending three or more concepts, our approach emphasizes selecting two main concepts to merge into a primary subject, while the remaining concepts are incorporated sequentially as secondary elements in the image. By considering similarity and sentiment scores, we strategically select objects and attributes for these additional concepts



**Figure 12: The results generated by our approach when processing more than two concepts. The outcomes presented are in line with the topics explored in *I Spy a Metaphor* [8].**

based on similarity and sentiment scores to create the final blended composition. Previous work, such as *I Spy a Metaphor* [8], also explores metaphorical representations. Unlike its method, which treats all concepts equally, our approach distinguishes itself by emphasizing the blending of specific concepts with metaphorically related objects. This not only facilitates creative exploration but also enhances the visual representation of abstract ideas [33]. Figure 12 demonstrates four topics that highlight the differences between our method and the previous approach. In the expression “Books are the mirror to the soul”, our system merges “books” and “mirror” into a primary subject, strengthening the metaphorical connection and establishing it as the main focus of the image. Additionally, we incorporate the “phoenix” element to represent the “soul”, showcasing the remaining concepts within the expression. In terms of the final output, the previous method relied on visual elaboration without clear semantic mapping, leading to scattered visual messaging. Our method, however, prioritizes element integration while preserving individual characteristics, making the intended message more easily recognizable by the audience.

## 7.3 Design with Generative AI

During our experiment, we found that the general T2I models do not always reflect users’ intent from their generated results. Participants commented that “sometimes it is tough to keep the AI model on track” (P11, Male, 27). In some circumstances, users “... have no confidence or idea of how the AI will interpret what (they) say”, and “...each prompt is a bit of a gamble” (P15, Male, 23). These responses revealed two major issues in designing with generative AI.

One is that natural language can be limited in its ability to convey certain design intentions. The second issue is that the capabilities of generative AI can sometimes be confusing to users due to their unclear boundaries.

The conversational style of current generative AI interfaces has catalyzed a novel trend in interaction design, enabling more intuitive communication with machines through natural language [34, 53]. However, the lessons we learned have revealed the limitations of natural language in conveying abstract concepts and representing tacit human knowledge, which is often learned through experience rather than explicit instruction. While natural language interfaces work well in domains where user intent is easily articulated, truly versatile AI interactions require systems that can understand abstract intent and respond to subtle cues, much like human communication. To mitigate this limitation, specialized tools can tailor the user interface to suit the needs of target users better. Given the potential shortcomings of natural language interfaces in handling multimedia content [46], specialized tools can effectively leverage alternative user interfaces (e.g., graphical user interfaces, or tangible user interfaces) to enhance user support for interactive tasks.

While generative AI offers the potential for processing multimodal information, it currently requires advancements in task decomposition and domain-specific expertise to address complex problems effectively. The evolution of software development shows that specialized techniques are needed to harness the unique capabilities of diverse applications. Despite the potential to integrate various information sources, current generative AI models sometimes struggle to produce coherent and reliable outputs, frequently generating unrealistic or nonsensical results. For example, Figure 8 illustrates how P6's result in T1 exhibits a visual discrepancy, such as an improperly lit cigarette. While Creative Blends focuses on design ideation, resolving such visual issues falls outside its scope. These common failures, often due to training data bias or model misalignment, can be mitigated through techniques such as retrieval-augmented generation or fine-tuning [29], though these methods require technical expertise and additional resources [20]. From a human-computer interaction perspective, challenges such as ambiguous input, lack of context, and inadequate instructions hinder users' ability to leverage generative AI for complex tasks. To improve usability, we propose using prompt engineering to break tasks into smaller sub-tasks and employing a guided task decomposition interface tailored to specific tasks, ultimately enhancing user understanding and problem-solving efficiency.

## 7.4 Limitations and Future Work

Our research is subject to several limitations that we intend to address in future work. Firstly, our evaluation was conducted using two design ideation tasks, each limited to under 20 minutes. In contrast, real-world design processes often span longer durations and encompass multiple iterative rounds of refinement and exploration. As a result, our analysis of user interactions and behaviors in creating visual blends was deliberately constrained to the ideation stage, reflecting the system's primary focus. In our evaluation, we employed a baseline consisting of Google Search and ChatGPT (GPT-3.5 with DALL-E 3) for comparison, aiming to

balance technical sophistication due to the automatic generation capabilities of AI models. Other comparisons might involve existing ideation methods like moodboarding [26] or mind-mapping [25] to investigate strengths and weaknesses further. Future work can also integrate Creative Blends into real-world design cases to observe how the system can facilitate designers' visual blend design processes.

Secondly, our system is designed primarily to inspire users during the ideation process, which currently limits the flexibility of the provided prompts for re-editing. Moreover, if used to generate the final result, the system still requires multiple trial-and-error iterations to achieve a satisfactory outcome for users. Future enhancements could include improving prompt editability to allow professional users greater flexibility in expressing their creativity, incorporating more control modalities to streamline the trial-and-error process, and extending support to the refinement stage further to enhance the system's utility and creative capabilities.

Thirdly, since our research primarily focused on understanding the system's potential within the design ideation context, a comprehensive analysis of prompt quality and generated result attributes, such as style and layout, was not undertaken. Nonetheless, we recognize that the quality of user-supplied prompts directly influences the generated outcomes. Future investigations could delve deeper into how the user's knowledge or experience with AI can affect their experience in generating visual blends. Additionally, exploring the impact of varying prompt formats on AI behavior and the resulting visual blend outputs could provide valuable insights into optimizing the creative process.

## 8 Conclusion

This paper introduces Creative Blends, an AI-assisted system designed to enhance the process of visual blend ideation by leveraging metaphors. Our system utilizes LLMs and commonsense knowledge bases to explore objects and their associated attributes, forming metaphorical connections with abstract concepts. It offers the capability to automatically generate blending proposals based on user selections, facilitating rapid creative realization for verification through the T2I model. To evaluate the system, we conducted a user study involving 24 participants who had AI experience. The findings demonstrate that Creative Blends has the potential to enhance the creativity of the generated ideation results and enable the expression of abstract concepts more metaphorically. Additionally, this research offers insights into user preferences regarding visual blend design and potential future approaches for supporting design with generative AI.

## Acknowledgments

We thank all anonymous reviewers for their insightful feedback. We especially thank Elad Richardson for his valuable input. This research was supported in parts by ICFCRT (W2441020), Guangdong Basic and Applied Basic Research Foundation (2023B1515120026), Shenzhen Science and Technology Program (KQTD20210811090044 003, RCJC20200714114435012, 20231122121504001), NSFC (62472288), ISF (3441/21, 3611/21, 2203/24), and Scientific Development Funds from Shenzhen University.

## References

- [1] Arjun R. Akula, Brendan Driscoll, Pradyumna Narayana, Soravit Changpinyo, Zhiwei Jia, Suyash Damle, Garima Pruthi, Sugato Basu, Leonidas Guibas, William T. Freeman, Yuanzhen Li, and Varun Jampani. 2023. MetaCLUE: Towards Comprehensive Visual Metaphors Research. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 23201–23211. <https://doi.org/10.1109/CVPR52729.2023.02222>
- [2] Réka Benczes. 2009. Visual metonymy and framing in political communication. *Journalism and Mass Communication Quarterly* 86, 1 (2009), 777–797.
- [3] Anna M. Borghi, Samuel Shaki, and Martin H. Fischer. 2022. Abstract concepts: external influences, internal constraints, and methodological issues. *Psychological Research* 86, 8 (01 Nov 2022), 2370–2388. <https://doi.org/10.1007/s00426-022-01698-4>
- [4] John Brooke. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.
- [5] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 1877–1901. [https://proceedings.neurips.cc/paper\\_files/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf)
- [6] Kelsy Cabello-Solorzano, Isabela Ortigosa de Araujo, Marco Peña, Luis Correia, and Antonio J. Tallón-Ballesteros. 2023. The Impact of Data Normalization on the Accuracy of Machine Learning Algorithms: A Comparative Analysis. In *18th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2023)*, Pablo García Bringas, Hilde Pérez García, Francisco Javier Martínez de Pisón, Francisco Martínez Álvarez, Alicia Troncoso Lora, Álvaro Herrero, José Luis Calvo Rolle, Héctor Quintián, and Emilio Corchado (Eds.). Springer Nature Switzerland, Cham, 344–353.
- [7] John M. Carroll and John C. Thomas. 1982. Metaphor and the Cognitive Representation of Computing Systems. *IEEE Transactions on Systems, Man, and Cybernetics* 12, 2 (1982), 107–116. <https://doi.org/10.1109/TSMC.1982.4308795>
- [8] Tuhin Chakrabarty, Arkadiy Saakyan, Olivia Winn, Artemis Panagopoulou, Yue Yang, Marianna Apidianaki, and Smaranda Muresan. 2023. I Spy a Metaphor: Large Language Models and Diffusion Models Co-Creat Visual Metaphors. In *Findings of the Association for Computational Linguistics: ACL 2023*, Anna Rogers, Jordan Boyd-Graber, and Naoki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 7370–7388. <https://doi.org/10.18653/v1/2023.findings-acl.465>
- [9] Erin Cherry and Celine Latulipe. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Trans. Comput.-Hum. Interact.* 21, 4, Article 21 (jun 2014), 25 pages. <https://doi.org/10.1145/2617588>
- [10] Lydia B Chilton, Ecezan Jen Ozmen, Sam H Ross, and Vivian Liu. 2021. VisiFit: Structuring Iterative Improvement for Novice Designers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (Yokohama, Japan) (CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 574, 14 pages. <https://doi.org/10.1145/3411764.3445089>
- [11] Lydia B. Chilton, Savvas Petridis, and Maneesh Agrawala. 2019. VisiBlends: A Flexible Workflow for Visual Blends. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland UK) (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3290605.3300402>
- [12] DaEun Choi, Sumin Hong, Jeongeon Park, John Joon Young Chung, and Juho Kim. 2024. CreativeConnect: Supporting Reference Recombination for Graphic Design Ideation with Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1055, 25 pages. <https://doi.org/10.1145/3613904.3642794>
- [13] João M. Cunha, Nuno Lourenço, Pedro Martins, and Penousal Machado. 2020. Visual Blending for Concept Representation: A Case Study on Emoji Generation. *New Gen. Comput.* 38, 4 (Nov. 2020), 739–771. <https://doi.org/10.1007/s00354-020-00107-x>
- [14] João Miguel Cunha, Pedro Martins, and Penousal Machado. 2020. Let's Figure This Out: A Roadmap for Visual Conceptual Blending. In *International Conference on Innovative Computing and Cloud Computing (ICCC)*. Association for Computational Creativity, University of Coimbra, Portugal, 445–452. <https://api.semanticscholar.org/CorpusID:228097521>
- [15] Pedro Miguel Amaral Melo da Cruz. 2016. *Semantic figurative metaphors in information visualization*. Ph.D. Dissertation. Universidade de Coimbra (Portugal).
- [16] Reinders Duit. 1991. On the Role of Analogies and Metaphors in Learning Science. *Science Education* 75, 6 (1991), 649–672.
- [17] Claudia Eckert and Martin Stacey. 2000. Sources of inspiration: a language of design. *Design Studies* 21, 5 (2000), 523–538. [https://doi.org/10.1016/S0142-694X\(00\)00022-3](https://doi.org/10.1016/S0142-694X(00)00022-3)
- [18] Charles Forceville. 1994. Pictorial Metaphor in Advertisements. *Metaphor and Symbolic Activity* 9, 1 (1994), 1–29. [https://doi.org/10.1207/s15327868ms0901\\_1](https://doi.org/10.1207/s15327868ms0901_1)
- [19] Giorgio Franceschelli and Mirco Musolesi. 2024. Creativity and Machine Learning: A Survey. *ACM Comput. Surv.* 56, 11, Article 283 (June 2024), 41 pages. <https://doi.org/10.1145/3664595>
- [20] Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Meng Wang, and Haofen Wang. 2024. Retrieval-Augmented Generation for Large Language Models: A Survey. arXiv:2312.10997 [cs.CL] <https://arxiv.org/abs/2312.10997>
- [21] Songwei Ge and Devi Parikh. 2021. Visual Conceptual Blending with Large-scale Language and Vision Models. In *International Conference on Innovative Computing and Cloud Computing (ICCC)*. Association for Computational Creativity. <https://arxiv.org/abs/2106.14127>
- [22] Joseph Grady, Todd Oakley, and Seana Coulson. 1999. Blending and metaphor. *AMSTERDAM STUDIES IN THE THEORY AND HISTORY OF LINGUISTIC SCIENCE SERIES 4* (1999), 101–124.
- [23] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139–183. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- [24] Angel Hsing-Chi Hwang. 2022. Too Late to be Creative? AI-Empowered Tools in Creative Processes. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (New Orleans, LA, USA) (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 38, 9 pages. <https://doi.org/10.1145/3491101.3503549>
- [25] Youwen Kang, Zhida Sun, Sitong Wang, Zeyu Huang, Ziming Wu, and Xiaojuan Ma. 2021. MetaMap: Supporting Visual Metaphor Ideation through Multi-dimensional Example-based Exploration. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (Yokohama, Japan) (CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 427, 15 pages. <https://doi.org/10.1145/3411764.3445325>
- [26] Janin Koch, Andrés Lucero, Lena Hegemann, and Antti Oulasvirta. 2019. May AI? Design Ideation with Cooperative Contextual Bandits. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland UK) (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3290605.3300863>
- [27] George Lakoff and Mark Johnson. 2008. *Metaphors we Live by*. University of Chicago Press, Chicago, IL.
- [28] Jiayi Liao, Xu Chen, Qiang Fu, Lun Du, Xiangnan He, Xiang Wang, Shi Han, and Dongmei Zhang. 2024. Text-to-Image Generation for Abstract Concepts. *Proceedings of the AAAI Conference on Artificial Intelligence* 38, 4 (Mar. 2024), 3360–3368. <https://doi.org/10.1609/aaai.v38i4.28122>
- [29] Hanchao Liu, Wenyuan Xue, Yifei Chen, Dapeng Chen, Xiutian Zhao, Ke Wang, Liping Hou, Rongjun Li, and Wei Peng. 2024. A survey on hallucination in large vision-language models. arXiv preprint arXiv:2402.00253 (2024).
- [30] Vivian Liu, Han Qiao, and Lydia Chilton. 2022. Opal: Multimodal Image Generation for News Illustration. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology (Bend, OR, USA) (UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 73, 17 pages. <https://doi.org/10.1145/3526113.3545621>
- [31] Leo S. Lo. 2023. The Art and Science of Prompt Engineering: A New Literacy in the Information Age. *Internet Reference Services Quarterly* 27, 4 (2023), 203–210. <https://doi.org/10.1080/10875301.2023.2227621>
- [32] Charlotte P Malycha and Günter W Maier. 2017. Enhancing creativity on different complexity levels by eliciting mental models. *Psychology of Aesthetics, Creativity, and the Arts* 11, 2 (2017), 187.
- [33] Edward F. McQuarrie and David Glen Mick. 1999. Visual Rhetoric in Advertising: Text-Interpretive, Experimental, and Reader-Response Analyses. *Journal of Consumer Research* 26, 1 (1999), 37–54. <http://www.jstor.org/stable/10.1086/209549>
- [34] Diego Moreira da Rosa, Leandro Soares Guedes, Monica Landoni, and Milene Silveira. 2023. Human Languages in HCI: Beyond User Interface Localization. In *HCI International 2023 – Late Breaking Papers*, Masaaki Kuroso, Ayako Hashizume, Aaron Marcus, Elizabeth Rosenzweig, Marcelo M. Soares, Don Harris, Wen-Chin Li, Dylan D. Schmorrow, Cali M. Fidopiastis, and Pei-Luen Patrick Rau (Eds.). Springer Nature Switzerland, Cham, 564–574.
- [35] Xiaohan Peng, Janin Koch, and Wendy E. Mackay. 2024. DesignPrompt: Using Multimodal Interaction for Design Exploration with Generative AI. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference (Copenhagen, Denmark) (DIS '24)*. Association for Computing Machinery, New York, NY, USA, 804–818. <https://doi.org/10.1145/3643834.3661588>
- [36] Matthew O. Peterson. 2019. Aspects of visual metaphor: an operational typology of visual rhetoric for research in advertising. *International Journal of Advertising* 38, 1 (2019), 67–96. <https://doi.org/10.1080/02650487.2018.1447760>

- [37] OpenAI Platform. 2024. Prompt engineering - OpenAI API. <https://platform.openai.com/docs/guides/prompt-engineering>. (Accessed on 11/11/2024).
- [38] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. In *Proceedings of the 38th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 139)*, Marina Meila and Tong Zhang (Eds.). PMLR, 8748–8763. <https://proceedings.mlr.press/v139/radford21a.html>
- [39] Elad Richardson, Yuval Alaluf, Ali Mahdavi-Amiri, and Daniel Cohen-Or. 2024. pOps: Photo-Inspired Diffusion Operators. arXiv:2406.01300 [cs.CV] <https://arxiv.org/abs/2406.01300>
- [40] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2020. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv:1910.01108 [cs.CL] <https://arxiv.org/abs/1910.01108>
- [41] Linda M. Scott. 1994. Images in Advertising: The Need for a Theory of Visual Rhetoric. *Journal of Consumer Research* 21, 2 (09 1994), 252–273. <https://doi.org/10.1086/209396> arXiv:<https://academic.oup.com/jcr/article-pdf/21/2/252/5093694/21-2-252.pdf>
- [42] Kihoon Son, DaEun Choi, Tae Soo Kim, Young-Ho Kim, and Juho Kim. 2024. GenQuery: Supporting Expressive Visual Search with Generative Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 180, 19 pages. <https://doi.org/10.1145/3613904.3642847>
- [43] Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. ConceptNet 5.5: an open multilingual graph of general knowledge. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (San Francisco, California, USA) (AAAI'17)*. AAAI Press, 4444–4451.
- [44] Zhida Sun, Sitong Wang, Wenjie Yang, Onur Yürüten, Chuhan Shi, and Xiaojuan Ma. 2020. "A Postcard from Your Food Journey in the Past": Promoting Self-Reflection on Social Food Posting. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference (Eindhoven, Netherlands) (DIS '20)*. Association for Computing Machinery, New York, NY, USA, 1819–1832. <https://doi.org/10.1145/3357236.3395475>
- [45] Alvaro Videla. 2017. Metaphors We Compute By. *Commun. ACM* 60, 10 (Sept. 2017), 42–45. <https://doi.org/10.1145/3106625>
- [46] Henrik Voigt, Monique Meuschke, Kai Lawonn, and Sina Zarrieß. 2021. Challenges in Designing Natural Language Interfaces for Complex Visual Models. In *Proceedings of the First Workshop on Bridging Human–Computer Interaction and Natural Language Processing*, Su Lin Blodgett, Michael Madaio, Brendan O'Connor, Hanna Wallach, and Qian Yang (Eds.). Association for Computational Linguistics, Online, 66–73. <https://aclanthology.org/2021.hcinlp-1.11>
- [47] Qian Wan, Xin Feng, Yining Bei, Zhiqi Gao, and Zhicong Lu. 2024. Metamorpheus: Interactive, Affective, and Creative Dream Narration Through Metaphorical Visual Storytelling. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 166, 16 pages. <https://doi.org/10.1145/3613904.3642410>
- [48] Qian Wan and Zhicong Lu. 2023. GANCollage: A GAN-Driven Digital Mood Board to Facilitate Ideation in Creativity Support. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference (Pittsburgh, PA, USA) (DIS '23)*. Association for Computing Machinery, New York, NY, USA, 136–146. <https://doi.org/10.1145/3563657.3596072>
- [49] Hung-Hsiang Wang and Jung-Hsuan Chan. 2011. An Approach to Measuring Metaphoricity of Creative Design. In *Design Creativity 2010*, Toshiharu Taura and Yukari Nagai (Eds.). Springer London, London, 89–96.
- [50] Sitong Wang, Savvas Petridis, Taeahn Kwon, Xiaojuan Ma, and Lydia B Chilton. 2023. PopBlends: Strategies for Conceptual Blending with Large Language Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (Hamburg, Germany) (CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 435, 19 pages. <https://doi.org/10.1145/3544548.3580948>
- [51] Yunlong Wang, Shuyuan Shen, and Brian Y Lim. 2023. RePrompt: Automatic Prompt Editing to Refine AI-Generative Art Towards Precise Expressions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (Hamburg, Germany) (CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 22, 29 pages. <https://doi.org/10.1145/3544548.3581402>
- [52] Zhijie Wang, Yuheng Huang, Da Song, Lei Ma, and Tianyi Zhang. 2024. PromptCharm: Text-to-Image Generation through Multi-modal Prompting and Refinement. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 185, 21 pages. <https://doi.org/10.1145/3613904.3642803>
- [53] Syed Mekaël Wasti, Ken Q. Pu, and Ali Neshati. 2024. Large Language User Interfaces: Voice Interactive User Interfaces Powered by LLMs. In *Intelligent Systems and Applications*, Kohei Arai (Ed.). Springer Nature Switzerland, Cham, 639–655.
- [54] Ping Xiao and Simo Linkola. 2015. Vismantic: Meaning-making with Images.. In *Proceedings of the Sixth International Conference on Computational Creativity (Park City, Utah, USA)*. 158–165 pages.