ProductMeta: An Interactive System for Metaphorical Product Design Ideation with Multimodal Large Language Models

ANONYMOUS AUTHOR(S)*†

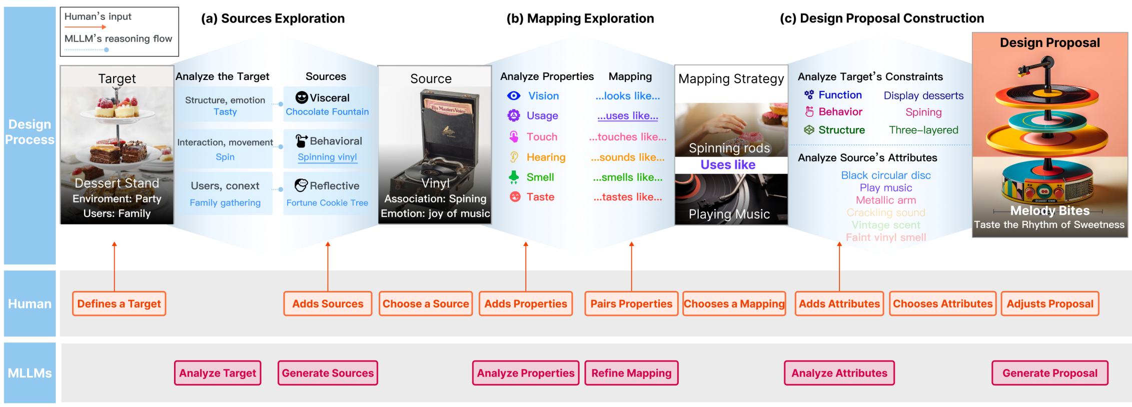

Fig. 1. An example of ProductMeta supporting novice designers in creating a metaphorical product design. (a) The system generates source inspirations from three perspectives: visceral, behavioral, and reflective for designers to choose from. (b) ProductMeta breaks down both the target and source into multi-sensory properties, enabling designers to recombine mapping ideas. (c) ProductMeta analyzes target constraints and source attributes, providing designers with detailed insights to address practical concerns and generate diverse proposals. Designers can freely navigation between modules.

Product metaphors, which involve creating products that convey meaning through metaphorical associations, are a powerful tool in product design. However, according to our formative study, novice designers often struggle to establish coherent links between target and source, to manage the complexity of diverse mapping possibilities and to balance product usability with metaphorical expression. To address these challenges, we introduce ProductMeta, a creative support tool designed to support novice designers in exploring and developing metaphorical product design. ProductMeta incorporates domain knowledge and decomposes the design process into iterative modules and framework-based interface, fostering both divergent and convergent thinking. Through user studies, we demonstrate that ProductMeta enables novice designers to generate diverse and contextually relevant design ideas by facilitating structured exploration. We conclude with design implications for human-AI co-creation.

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

Additional Key Words and Phrases: Creativity support tool, Product Design Ideation, Metaphor Design, Machine Learning

---

Manuscript submitted to ACM

ACM Reference Format:
Anonymous Author(s). 2018. ProductMeta: An Interactive System for Metaphorical Product Design Ideation with Multimodal Large Language Models. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 34 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Metaphors are essential tools in product design to foster creativity and innovation[47, 58, 91] which have become key differentiators for product success [31, 43, 68] . Metaphorical product design involves mapping properties from a *source* (a distinct entity) to a *target* (the product) to bridge abstract ideas and tangible forms [36, 39]. This process helps to enhance the product's novelty [47], evoke emotions [62, 68], improve user experience [29, 31], and encourage post-use reflection [59, 64].

A classic example is the Anna G. bottle opener, which incorporates the image of a woman in its design [4] (Fig. 2 .a). The product's elegant form evokes beauty, while the rising arms of the opener signal progress and joy, aligning the act of opening a bottle with a sense of accomplishment. This design also fosters a reflective experience, drawing attention to gender issues, potentially provoking thoughts around sexual harassment [59, 64]. This integration of visceral, behavioral, and reflective elements exemplifies how metaphors in product design can engage users on multiple levels [68]. Additional examples are shown in Figure. 2.

Despite its potential, metaphorical product design remains challenging for novice designers [16, 29]. The process typically involves first identifying a source that aligns with the intended meaning and then creating a physical mapping to convey that meaning [28, 29, 45, 46]. Crafting effective product metaphors requires not only associative thinking but also complex mapping techniques [46, 62, 91]. While extensive research exists on metaphors in fields like visual design [24, 54] and architectural design [16, 17], there remains a notable knowledge gap in product design. To better address this gap and understand the challenges faced by novice designers, we conducted a formative study with 10 product designers to examine their ideation process and identify areas for improvement.

The study revealed three key challenges in designing product metaphors: (1) finding creative, contextually relevant source; (2) exploring diverse mapping possibilities; and (3) balancing functional constraints to maintain feasibility and meet physical requirements when developing proposals. Overall, designers struggle to fully explore relevant ideas while ensuring their designs are both innovative and practical.

Traditional creative support tools, like mood boards [54] and metaphor cards [60, 61] rely heavily on the designers' cognitive ability to manually establish connections. Database-driven tools [6, 33, 65, 73, 82] also fall short, offering limited data and lacking context-relevant metaphors necessary for product design. Visual blend tools [24, 25] rely on image collages and literal visualizations, are unsuitable for product design due to the need for abstraction and simplification before blending. While the emerging capabilities of Multimodal Large Language Models (MLLMs) demonstrate potential in creative thinking [38], multimodal analysis [14], and image generation [48, 77, 94], they remain inadequate for supporting novice designers. First, MLLM-based tools often rely on a one-shot generation process, which limits the ability of novice designers to further explore creative alternatives or iteratively refine a specific proposal [84]. Second, their lack of domain-specific design knowledge restricts the generation of contextually relevant ideas. Consequently, this limitation hinders the development of a systematic metaphor ideation process.

We developed **ProductMeta**, a creativity support tool (CST) that integrates domain knowledge and framework-based interaction to address 3 key challenges in metaphorical product design. Powered by an instruction-tuned GPT-4o model with few-shot prompt engineering, ProductMeta helps novice designers to ideate, create, and iterate on metaphorical products through human-AI collaboration. ProductMeta suggests target-source pairs from various associative perspectives to inspire creative connections. Once a source is chosen by the designer, it breaks down both the target and source into key properties, and uses an interactive prompt panel to enable multi-sensory mapping explorations. During design proposal construction, ProductMeta analyzes the target functions and offers flexibility to refine ideas while ensuring alignment with practical constraints. ProductMeta allows designers to freely navigate between modules, promoting a **non-linear and controllable** workflow that encourages exploration and iterative refinement.

We conducted a within-subjects study with 16 novice designers to evaluate ProductMeta in terms of metaphor and creative process support, system usability, and design output quality. To understand how ProductMeta improves upon existing systems, we compared it to the GPT-4o-powered ChatGPT as a design tool. Each participant completed two separate and counterbalanced creative product metaphor tasks, with the target product selected from a predefined list. Results showed that ProductMeta provided superior support across all three stages of metaphor creation: source exploration, mapping exploration, and design proposal construction. Compared to GPT-4o, ProductMeta offered a more supportive, engaging and expressive experience, enhancing exploration while reducing perceived effort. Final outputs were evaluated by four design experts. Qualitative analysis revealed that ProductMeta's designs were of higher quality due to more context-specific source selections and the use of appropriate mapping strategies. Finding suggests potential integration of natural language interaction with structured framework-based interfaces to create a more fluid and flexible design experience. Our contributions include:

- Challenges and difficulties encountered by designers when creating product metaphors, identified through a formative study.
- ProductMeta, an interactive system designed to support diverse idea exploration while considering physical requirements and functional constraints in metaphorical product design. It enables a non-linear, controllable design workflow.
- An evaluation of ProductMeta's efficacy was conducted through a user study with 16 novice designers, providing an in-depth understanding of how the system facilitates idea exploration, iteration, and refinement throughout the design process.

2 RELATED WORK

Metaphors are vital for fostering creativity in design, enabling connections between disparate domains and facilitating innovation [17, 47, 58, 75, 76, 80]. In interface [12, 35, 49, 76], architecture [17, 19, 67], and visual design [3, 79], metaphors help designers understand complex concepts and reframe problems [18, 46, 58, 78]. In product design particularly, metaphors bridge abstract ideas and tangible forms, enhancing novelty, emotion, and user experience [46, 91].

2.1 Traditional Methods for Metaphorical Product Design

Classic methods, such as SCAMPER, TRIZ, and tools like brainstorming cards and empathy maps, support source identification and idea organization [34, 50, 61, 81]. Cognitive strategies like analogical reasoning [39, 46] and conceptual blending [32, 47] guide mapping, while frameworks like Function-Behavior-Structure ensure feasibility. These methods rely heavily on experience, posing challenges for novices who struggle with systematic ideation and the integration of diverse techniques.

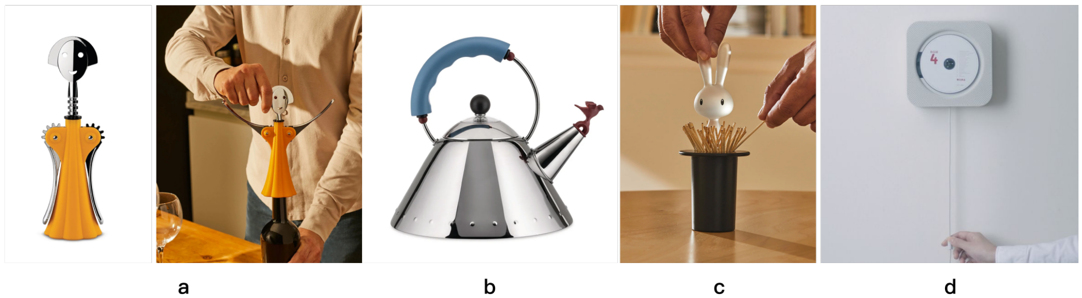

Fig. 2. Examples of classic metaphorical product design: (a) *Anna G. Wine Opener*: incorporates a female figure (source) with a bottle opener (target) by mapping the figure's rising arms (source's property) to the corkscrew's opening action (target's property), creating a playful yet controversial interaction.; (b) *9093 Kettle*: associates a bird-shaped whistle (source) with a kettle (target) by mapping the bird's figure (source's property) to the kettle's spout (target's property), producing a delightful, joyful interaction; (c) *Magic Bunny Toothpick Holder*: connects pulling a toothpick (target) with pulling a rabbit from a magician's hat (source), evoking playful and humorous emotions; (d) *Naoto's Wall-Mounted CD Player*: links pulling an old exhaust fan (source) to opening a CD player (target), introducing a more intuitive user interaction.

2.2 Computational Tools for Metaphor Creation

2.2.1 Non-generative Tools. Existing tools for metaphorical product design generally fall into two categories: source-target identification and mapping assistance. Tools such as metaphor databases (e.g., Metaphor Map and ConceptNet [6, 82]) and visualizations (e.g., Sankey Diagrams [42]) help designers identify potential source-target pairs. However, tools like MetaMap, supporting visual metaphor ideation, rely on predefined datasets (e.g., the Small World Project), which can limit the diversity of sources they recommend [54].

Techniques like image fusion and semantic decomposition further aid mapping [25, 52]. While mapping tools like VisiBlends can help visualize the relationship between source and target at a superficial level, they fail to seamlessly and intuitively blend metaphorical components into a coherent product design. Also, these tools address only specific stages of the design process.

2.2.2 Generative Tools. Recent advances in multimodal large language models (MLLMs) have shown potential in creative domains, enabling tasks such as metaphor understanding (e.g., MetaCLUE [3], metaphor in software engineering [51]), metaphor exploration (e.g., Metaphorian [55], visual metaphor creation [20]), and visual blending (e.g., PopBlends [92]). While these tools excel in ideation, they fall short in metaphorical product design, which requires embodied understanding, multi-sensory considerations, and practical feasibility. Designing metaphorical products involves distilling source and target concepts into key design elements—such as structures, shapes, or textures—while ensuring usability and feasibility. Current tools often lack domain-specific knowledge, limiting their effectiveness for novice designers.

2.3 Challenges in Human-AI Collaboration

MLLMs enable new possibilities for human-AI collaboration but face challenges in creative workflows. Linear interaction models, such as those in ChatGPT, hinder iterative and exploratory design processes, requiring structured interfaces to support non-linear backtracking and refinement [53]. Tools like Graphologue and generative.fashion show promise in addressing these issues by introducing graphical dialogues and iterative design interfaces [30, 53].

Another challenge is that users often struggle with crafting effective prompts, leading to suboptimal outcomes and a sense of lost control [15, 95]. This issue stems from end-to-end nature of generative models, where the absence of intermediate processes makes it hard to control outputs solely through textual prompts [63]. Decomposition frameworks, such as Luminate and Prompting for Discovery, mitigate this by breaking down prompts or visualizing intermediate outputs to guide iterative exploration [5, 83].

Faced with these challenges, our work focuses on supporting metaphorical product design space exploration.

3 FORMATIVE STUDY

To understand the design process, challenges, and anticipated support needs in metaphorical product design, we conducted semi-structured interviews with 10 designers (2 male, 8 female, P1-P10) recruited through a university platform. Participants' mean age was 29.07, and the SD was 4.74 (M: Mean = 30.53, SD = 5.42, F: Mean = 27.6, SD = 3.54). All participants had prior experience in metaphorical product design, spanning novice to expert levels, distinguished by years of professional training in product design. Participant demographics are detailed in Section A.1. Each interview was conducted in person, lasting approximately 40 minutes, and participants were compensated with $15. The interviews explored four themes: (1) general processes used in creating metaphors, (2) challenges faced during the process, (3) desired support tools for metaphorical product design, and (4) current limitations of MLLM-based tools in supporting metaphor creation.

3.1 Analysis

Interview audio recordings were manually transcribed and analyzed using thematic analysis. One author initially created a codebook for all transcripts through an inductive approach, which was then discussed and refined by two authors. After reaching a consensus on the codebook, two randomly selected transcripts were jointly coded to resolve discrepancies and finalize the codebook. The remaining transcripts were independently coded by the two authors, followed by collaborative meetings to refine interpretations and adjust the coded data.

3.2 Workflow of Product Metaphors

Based on participant responses and a literature review, we identified three main phases in the metaphor creation process:

(1) **Identifying Associations and Sources:** Designers first brainstorm the target's properties from various perspectives (e.g., a kettle whistles when boiling). Based on these properties, they then explore creative associations between the target and potential sources to form a target-source pair (e.g., the whistle of a kettle resembles a bird's whistle, as illustrated in Figure 2. b).

(2) **Mapping Source Properties to the Target:** Designers map the properties of the chosen source to the target, ensuring the metaphor is identifiable and conveys the source's meaning (e.g., designing the kettle's spout as a bird-shaped whistle activates when steam passes through).

(3) **Constructing the Design Proposal:** Designers visualize the concept and refine details like color, texture, and form to enhance the metaphor's expression. (e.g., using red to accentuate the bird-shaped whistle).

Creating product metaphors is not a linear process but rather a dynamic and iterative one, involving frequent shifts between divergent and convergent thinking. Designers often revisit earlier steps when ideas fail to meet expectations. As P3 noted, "It's about some A-ha! moment. There's no strict logic to it".

3.3 Challenges and Desired Support for Creating Product Metaphors

3.3.1 Lack of Ideas and Inspiration for Creative Sources. All participants expressed difficulty in generating creative and reasonable connections between the target and source. Effective connections require designers to understand the target's hidden qualities [78], followed by brainstorming potential sources and their connection with the target. Designers usually turn to mood board tools like Pinterest and Google Search for sources inspirations. However, these tools often fail to provide meaningful insights. As P4 noted, "I think the hardest part is making it reasonable. You can force a connection between anything, but the question is whether it makes sense." Novice designers particularly struggle with creative blocks, often generating ideas that are either overly conventional or far-fetched. This challenge is frequently linked to a limited understanding of the target or a lack of diverse life experiences. P1 observed, "Finding clever connection points requires careful observation of life." P7 highlighted, "Novice designers are often unfamiliar with cultural contexts or fail to conduct thorough preliminary research when starting a project".

3.3.2 Cognitive Load and Overlooked Properties in Mapping. After identifying a source, designers often experiment with various property mappings, such as form, behavior, and touch. As there are too many properties in considers, Designers are concerned that they may overlook non-salient properties and fail to explore sufficiently broad mapping strategies. As P1 noted, "How did I connect a face mask and tissue paper? ..... They are not only similar in shape but also in how they are used". Cognitive overloads frequently occurs due to the vast number of possible combinations within and between properties [11].

3.3.3 Balancing Trade-offs in Detailed Design. While conceptualizing metaphorical designs is relatively straightforward, designers spend significant time sketching and visualizing ideas, often navigating two key trade-offs:

Balancing Usability with Metaphorical Expression. Designers need to maintain the usability of the target product while incorporating metaphorical elements from the source. The integration of source attributes sometimes conflict with the functionality and structure requirement, which can diminish the usability of the target product and require designers to make adjustments:"We initially designed a coffee handle as a small triangle to reflect the shape of a penguin's beak, but testing showed it was hard to handle, so we made it larger" (P4).

Determining Implicit and Explicit Mapping Styles. The level of metaphor explicitness can vary greatly. During the visual design stage, designers create multiple sketches to decide which mapping style works best within the product's context. As P6 noted: "The depth of the metaphor should match the context. Not everything needs to be subtle, but the degree of metaphor should fit the design's application." This decision-making process is highly contextual, influenced by both the designer's intentions and the expected user experience.

3.4 Limitations of Collaborating with Current MLLMs

Five participants reported prior experience of using generative tools (including GPT-4o and Midjourney), and we asked them to identify the problems and limitations they encountered when co-creating with current AI tools. Three main issues with the current process:

3.4.1 Lack of Creativity in Source and Mapping Ideas: Participants observed that the AI-generated sources often lacked creativity. For instance, when designing an alarm clock, AI might suggest typical sources like the sound of water flow, something easily conceived by most designers. As P3 noted, "The ideas generated by AI were ones I had already dismissed during the initial brainstorming session." Additionally, during the mapping, the images generated by AI often produce overly superficial combinations, focusing on graphical or literal associations rather than deeper conceptual links. For example, the AI might literally map the graphic of a sunrise onto the dial of an alarm clock, neglecting more implicit and mapping, such as changing light patterns that mimic the gradual transition of dawn. The underlying problem identified by participants is that AI tools analyze targets and sources too superficially, lacking in-depth property analysis and scenario-based thinking (P1, P4, P9).

3.4.2 Lack of Designers' Control: Designers revealed that current AI tools provide direct results without allowing control over each step. P4 mentioned, "It should be a step-by-step process, and I can filter a little at each step, the chain should be transparent. GPT is hard to adjust the process in between steps." Additionally, they struggle to clearly articulate specific design properties, reducing the effectiveness of the control. (P4, P6)

3.4.3 Leaning to Convergent and Detail Adjustment in the Mapping Phase: During the mapping phase, AI-generated outputs often include highly detailed images, which encourages designers to focus on convergent thinking and spend excessive time refining these images rather than engaging in broader metaphor ideation [27, 84]. As a result, the detailed nature of AI outputs tends to limit exploration by narrowing the focus to specific visual elements. As P11 remarked, "I think before you show me the image, it needs to be semantically reasonable—make sure it sounds right to me first, and then ensure it looks right"

4 DESIGN GOALS

The formative study and content analysis revealed the need for assistance in source exploration, mapping exploration, and design proposal construction phases. To address these needs, we derive the following design goals (DGs) to guide the design of ProductMeta:

DG1. Provide Diverse Source Inspirations in Source Exploration The formative study revealed that designers prefer AI to suggest a range of inspirational ideas automatically during the source brainstorming phase, allowing designers to select from these options. The system should present diverse and creative sources intuitively. To ensure this diversity and creativity, we draw on Norman's three levels of emotional design—visceral, behavioral, and reflective—which facilitate identifying sources that resonate emotionally across different dimensions [68]. This framework is widely used in metaphorical product design [28, 64, 91]. Additionally, the system should enable designers to easily refine AI suggestions or introduce their own ideas.

DG2: Support Multi-sensory Exploration in Mapping Strategies. To assist designers in exploring diverse mapping strategies, our system should enable users to quickly discover, implement, and compare creative mapping approaches. Previous design research highlights that multi-sensory properties such as touch, sound, and interaction can enhance cognitive and emotional responses, fostering deeper conceptual understanding and stronger user-product connections [69]. Embodied representation is crucial for understanding metaphors in product design. [9, 57, 87]. Thus, the system should inspire multi-sensory mappings while maintaining the designer's creative flow. To align with designers' preference for conceptual exploration over highly detailed artwork during the early ideation phase, the system should encourage a broad exploration of mapping possibilities before progressing to the detailed design or image refinement.

DG3: Balance Usability and Metaphorical Expression in Design Proposal Construction. The system should ensure that the target's usability remians intact while integrating metaphorical elements from the source during image generation. Tools for analyzing product functionality should assist designers in identifying and preserving essential elements during the proposal phase. Additionally, the system should provide a wider range of design proposals, incorporating both target-driven and source-driven mappings, as well as different styles of metaphor expressions.

DG4: Support Shifts between Divergent and Convergent Ideation. The system should support smooth transitions between divergent and convergent thinking by enabling designers to zoom in on specific design proposals and zoom out to compare different sources or mapping strategies easily. To prevent designers from fixating on detail adjustments in AI-generated images too early, the system should help maintain a broader perspective, prioritizing metaphor ideation over fine-tuned image refinement.

5 PRODUCTMETA

ProductMeta is an interactive ideation tool that helps novice designers create metaphorical product designs through an exploratory and iterative process. Novice designers are defined as users who have received some design education but lack experience in metaphorical design tasks. For example, a visual communication student tasked with designing a commemorative item for their campus may find it challenging to integrate visual design principles with the nuances of physical product design.

5.1 System Overview

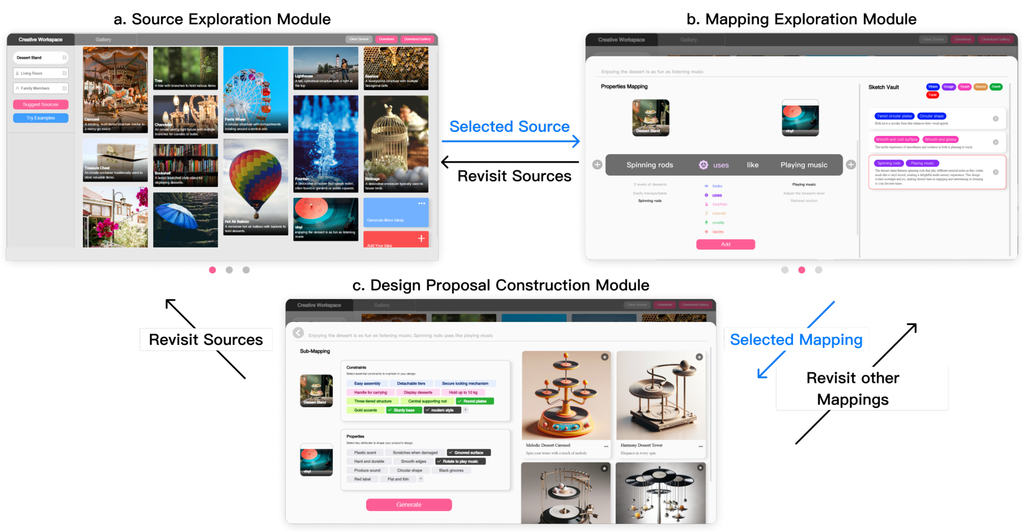

Fig. 3. User Interface of ProductMeta, consisting of (a) Source Exploration Module, (b) Mapping Exploration Module, and (c) Design Proposal Construction Module. These modules are interactive and synchronized, supporting a dynamic design process.

ProductMeta supports metaphorical product design through three core modules: 1) **Source Exploration** (Fig. 3. a), which allows designers to explore and identify potential sources; 2) **Mapping Exploration** (Fig. 3. b), where designers experiment with different property mappings between the source and target; and 3) **Design Proposal Construction** (Fig. 3. c), where designers refine their design details.

These modules are both interactive and synchronized: modifications in one module automatically update the others, ensuring a fluid and dynamic design process, as shown in (Fig. 3). The system provides a seamless experience across three interconnected pages, with a tab bar allowing users to navigate easily between modules. We split different modules in three pages, to reduce distractions and supports clarity [37], especially for novice users.

Designers are required to input design requirements, including the target product, usage environment and target user. The system supports a wide range of product types, including kitchenware, household items, and merchandise, allowing designers to input various specifications as needed.

We chose GPT-4o as the underlying model as it is a state-of-the-art MLLM that integrates essential capabilities for metaphorical product design. These capabilities include associative thinking, understanding of the physical world, contextual awareness, visual thinking, and both inductive and deductive reasoning [14, 38, 94]. GPT-4o's multi-modal understanding and output capabilities make it an ideal foundation for supporting design tasks.

5.2 Design of ProductMeta

5.2.1 Source Exploration.

Source Waterfall. The Source Waterfall displays sources in a vertical layout. Users input design requirements. The module generates suggestions in a full-screen, mood-board-like interface, allowing users to freely explore sources without distractions. To ensure diverse and relevant sources (DG1), ProductMeta analyzes the target's characteristics first. This analysis is crucial for generating insightful metaphors, as effective metaphors draw on the target's distinct attributes [28, 29]. For brainstorming metaphorical sources, we draw on Norman's three levels of emotional design: visceral, behavioral, and reflective [2, 68] (DG1). Specifically, the system analyzes the target's key characteristics across these levels:

(1) **Sensory Level:** Visceral characteristics, including visual, smell, taste, touch, sound.
(2) **Behavioral Level:** Characteristics tied to product interaction and movement.
(3) **Reflective Level:** Contextual characteristics, including the user's socio-cultural background, personal identity; as well as the cultural and emotional associations the product carries [8, 64].

To ensure a more comprehensive analysis, the system first decomposes the product into its constituent components and analyzes each component's sensory and behavioral characteristics [85].

Based on this analysis, the module generates a set of sources. Typically, the system generates 5 source options at a time, but this number can be adjusted to 10 based on user preference.

Metaphor Card. To address designers' struggle in establishing clear connections between the target and source (DG1), ProductMeta offers a metaphor card that provides a structured, intuitive way to present the association.

Each metaphor card consists of the following elements:

(1) **Source Name**
(2) **Association**: A sentence describing the connection between the target and the source.
(3) **Source Image**: An image associated with the source, retrieved from an image dataset to help designers intuitively understand the relationship through images. The dataset can be configured to pull images from either an external stock photo platform or a local image database. In our system, we utilized the Pexel[1] dataset. This dataset contains a large number of high-quality photographic images, which are manually screened and labeled . The source name is used as the search query, and a random image from the top-K (K<5) is returned each time.

[1]https://www.pexels.com

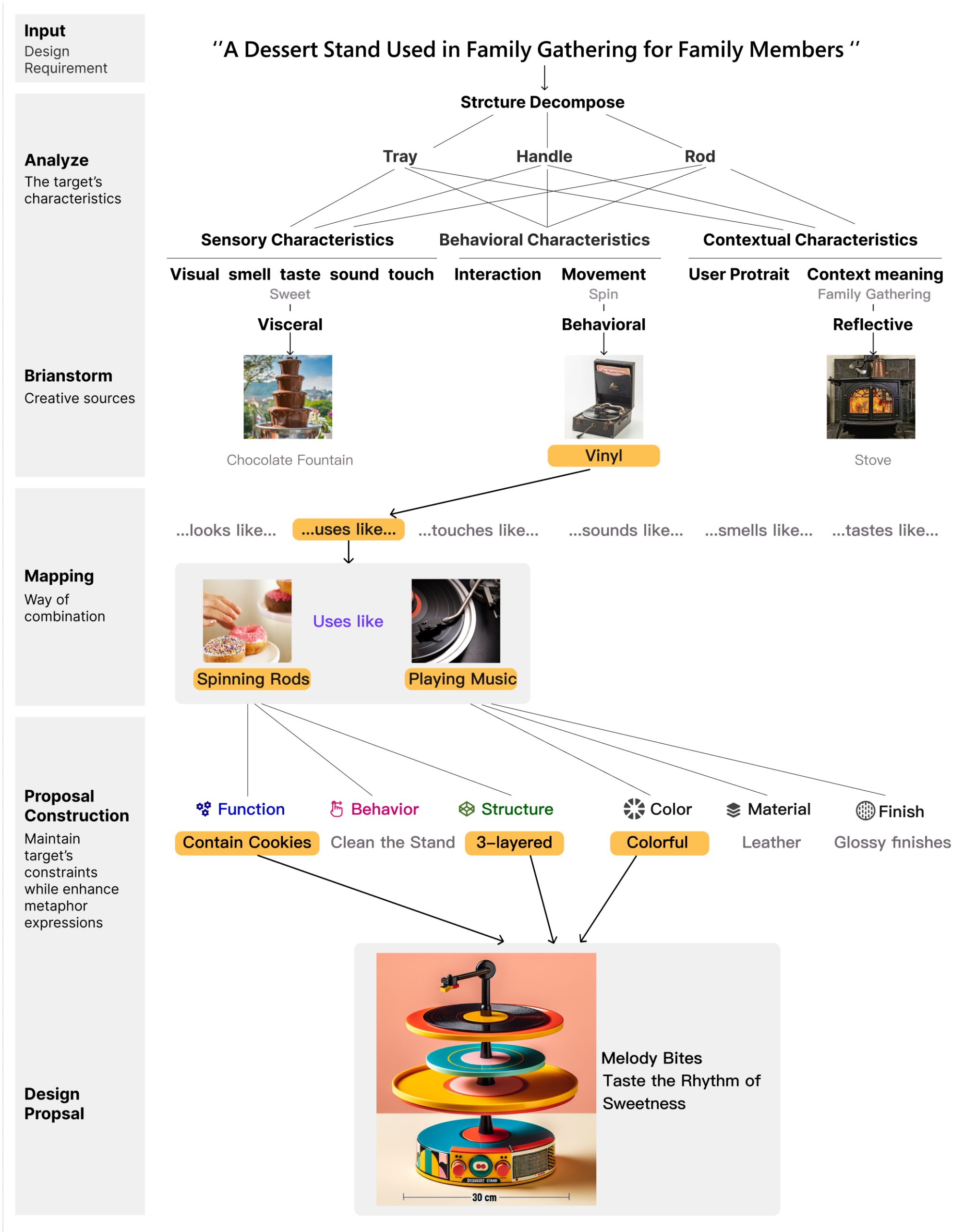

Fig. 4. An illustration of the ProductMeta workflow to design a dessert stand. The yellow blocks indicate the selections made by the user.

(4) **Additional context**: When hovering over the card, designers can see further description of this target-source pair: 1) Potential usage scenarios to deepen understanding of how the source relates to the design context; 2)

Emotional effects and the transferred meaning of the source; 3) Possible mapping strategies to envision the mapping next step. The description are generated by GPT-4o with a few-shot example.

To optimize screen space, the height of each metaphor card is capped to 150px, allowing at least 10 ideas to be visible on the screen at once. This height is sufficient to display the description on the back of the metaphor card.

5.2.2 *Mapping Exploration.*

Property Pairing Panel. The Property Pairing Panel was designed to facilitate multi-sensory mapping explorations. This panel consists of three interactive bars, each defining a part of the mapping strategy in the following structure: the target's **... property** [target's properties bar] + **in which way** [sensory attributes bar] + **like** + the source's **... property** [source's properties bar]. For example, "the spout looks like a bird." Designers can explore multiple mapping strategies by interacting with the panel.

Integrated with an MLLM component, the panel analyzes and suggests mappable properties from a multi-sensory perspective. Building upon the source-target pair selected in the Source Exploration Module, the MLLM analyzes the target and source images along with their associated descriptions to suggest properties. It generates a list of 12 properties, categorized into 6 sensory groups: Vision, Usage, Touch, Sound, Smell, and Taste. The categories are informed by existing metaphorical product design frameworks [12, 29, 62]. Each property is prompted to be less than 6 words and exclude the target and source names.

To ensure the system generates suitable properties for product design, we leverage few-shot learning. A dataset of 150 classic cases of metaphorical product design was curated, with each case analyzed by two professional product designers for the properties used in the mapping [2]. This dataset serves as the foundation for few-shot demonstrations, where two random cases are selected each time to increase diversity.

Concept Vault. Once users are intrigued by a mapping in the Property Pairing Panel, they can add it to the Concept Vault. The Concept Vault is designed to quickly refine a mapping concept into a more concrete design description. This allows designers to preview the concept in a more detailed way. MLLMs will explain how to merge these two attributes based on the mapping strategy chosen on Property Pairing Panel. MLLMs are utilized for generating connection descriptions because they excel at identifying links between seemingly unrelated properties, expanding design possibilities and inspiring new ideas.

The system does not incorporate image generation in this module, as it focuses on conceptual ideation and defining the metaphor concept. Our formative study found that introducing images too early can lead to excessive iteration on visual details, limiting creative exploration (Section 3.4.3).

Property Extractor. In product design, a source often includes various subtypes (e.g., different shapes or versions), and relying solely on system-suggested images may not align with the designer's vision. To better support this alignment, we encourage designers to find and upload source images from the web that more closely match their intended concept. Previous study revealed that designers struggle to translate mood boards' results into specific keyword inputs for AI systems [26, 72]. The Property Extractor solves this by enabling designers to upload images of source and target objects for property elicitation. The system then extracts relevant properties from these images and adds them to the property panel.

5.2.3 *Design Proposal Construction.*

[2]https://github.com/productmetaphor/Product-Metaphor-Dataset

Constraints Panel and Attributes Panel. Based on our formative study, designers make visual adjustment to balance usability and metaphor expression when constructing proposals (Section 3.3.3). While current MLLMs can generate high-quality product renderings, their ideas generated are infeasible [41, 74].

The following panels break down the design elements that designers need to consider in detail design. Designers can select keywords to guide which characteristics of the target they want to maintain or incorporate in the final design (DG3):

(1) Constraints Panel: The system analyzes the target's functional, behavioral, and structural elements using Gero's Function-Behavior-Structure (FBS) model [40].
(2) Attributes Panel: This panel focuses on the source's design details, sucha as form, color, texture, material properties [10, 66], ensuring that metaphorical elements are coherently integrated into the product.

To help designers easily navigate these elements, the system color-codes the different categories they need to consider.

Proposal Creation Panel. To broaden users' divergent thinking, this component generates distinct design proposals reflecting various mapping styles. Given previous design configurations, ProductMeta selects four mapping style keywords. The keywords includes Literal Mapping, Abstract Mapping, Source-Driven Mapping, Target-Driven Mapping as referenced in previous research [29, 46, 91], along with design styles (e.g., playful, elegant, modern, minimalist, poetic, humorous). Using these keywords, the component generates four distinct proposals, each including: 1) Name of the design; 2) A slogan; 3) Description: User scenario, metaphor connection, and the intended meaning and emotion; 4) product renderings.

5.3 User Scenario

The following scenario illustrates how Lisa, a junior product designer would use the ProductMeta to design a *dessert stand*. Lisa wants to explore how to make her design unique and innovative with the help of ProductMeta.

Exploring Sources. First, Lisa defines her design requirements in the Search Panel, specifying the product category as "dessert stand," the usage environment as "family party," and the target users as "family members" (Fig. 5. a). After clicking "Suggest Sources," the system presents five metaphorical sources in the form of Metaphor Cards. Inspired by these suggestions, Lisa clicks "Add Your Idea" and input the source she has come up with – a lotus, prompting the system to generate a Metaphor Card with a connection between the lotus and the dessert stand (Fig. 5. d). However, Lisa finds the association misaligned with her vision and manually edits it in the Source Addition Window (Fig. 5. d).

Next, Lisa explores other sources. A "vinyl" catches her attention. Hovering over it reveals how the "joy of listening to music" could be connected to the "joy of tasting sweets" (Fig. 5. c). Intrigued, Lisa clicks the vinyl Metaphor Card and is navigated to the Mapping Exploration Module.

Exploring Mappings. In the Property Pairing Panel, Lisa experiments with mapping properties from the "vinyl" to the "dessert stand" across multiple sensory dimensions (Fig. 6). As she interacts with the Attributes Bar, the system dynamically updates the properties for both the target and the source, allowing Lisa to explore different mapping strategies, such as vision, usage, touch, sound, smell, and taste.

Lisa suddenly recalls a digital vinyl, more aligned with her goal to engage a younger audience. To clarify her design intent, she uploads an image of this vinyl. The system automatically adds relevant properties to the Attributes Bar, expanding her options. In addition to the AI-suggest properties, Lisa can also add her own properties by clicking the "Plus" button in each side (Fig. 6. c).

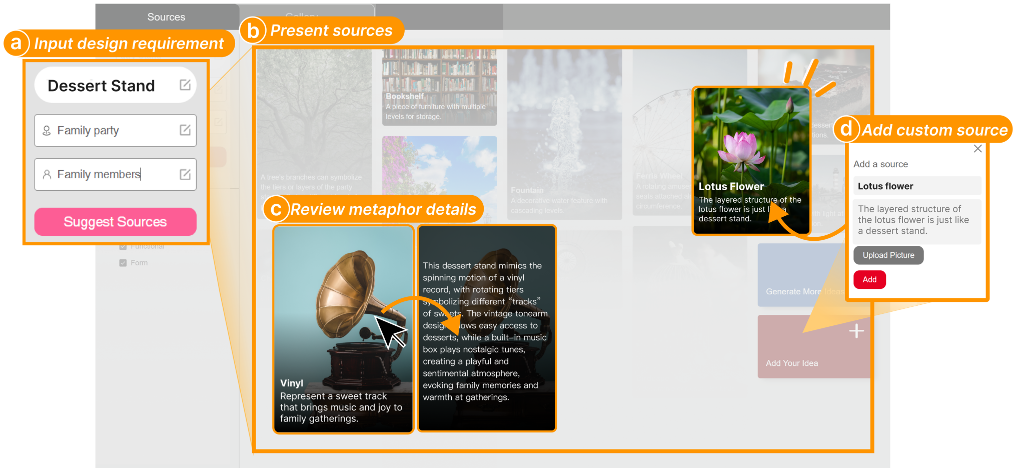

Fig. 5. The Source Exploration Module allows users to explore target-source pairs through four components: (a) Search Panel for specifying the target, (b) Source Waterfall displaying source inspirations, (c) Metaphor Card showing target-source relationships, and (d) Source Addition Window for manually adding sources and defining their associations. The yellow elements illustrate the flow of user interactions.

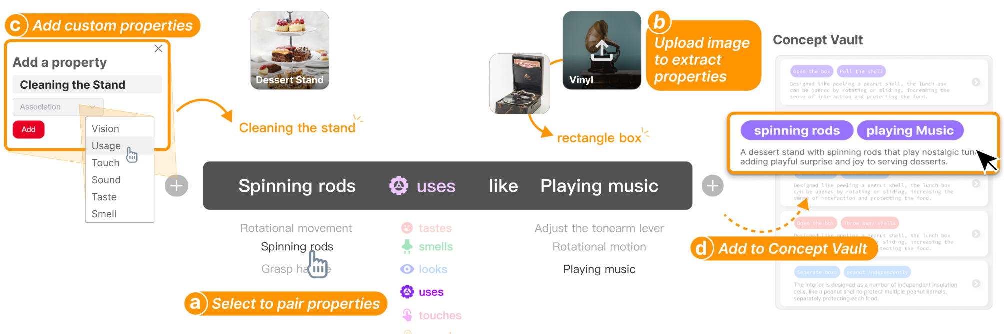

Fig. 6. The Property Mapping Module enables users to explore mapping strategies via: (a) Property Pairing Panel, (b) Property Extractor, allowing users to upload images for property elicitation, and (c) Sketch Vault, where users can preview, refine, and compare mapping strategies.The dashed lines represent the transmission of data.

After experimenting with various mappings, Lisa finds one compelling -spinning rods uses like playing music. She adds it to the "Concept Vault" (Fig. 6. d). AI refined this concept–features a dessert stand with spinning rods that play tunes. By saving it to the vault, Lisa can compare this mapping with other ideas she has collected (DG4). Lisa can then proceed to the Design Proposal Construction Module by selecting a mapping from the Concept Vault.

If Lisa does not find any promising mappings in the Concept Vault or wants to explore new ideas, she can return to the Source Exploration module to investigate other sources.

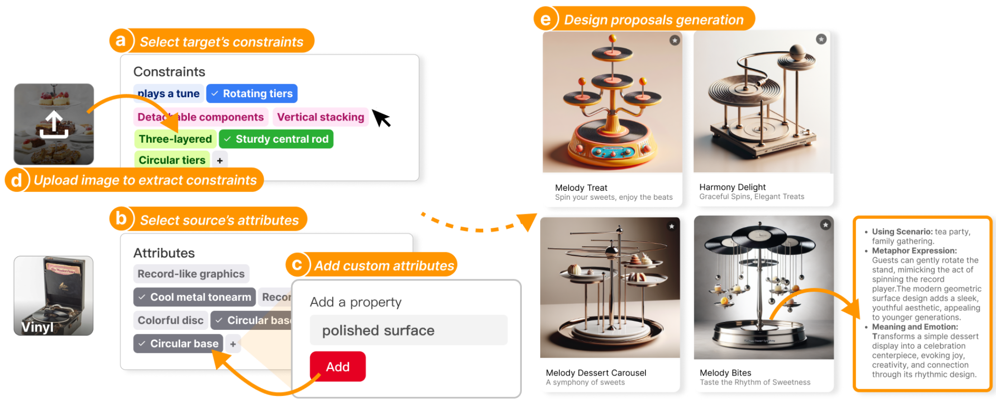

Fig. 7. The Proposal Construction Module enables users to refine design proposal. Key components include: (a) Constraints Panel, (b) Attributes Panel, (c) Addition Window for adding custom attributes, (d) Property Extractor for uploading images to elicit properties, (e) Proposal Creation Panel.

Constructing Design Proposals. In the Design Proposal Construction Module, Lisa refines the constraints for her design and generates design proposals by selecting the corresponding keywords in the Constraints Panel and the Attributes Panel. She maintains the sturdy central rod from the dessert stand while incorporating the vinyl's cool metal tonearm (DG3; Fig. 7. a,b). The color-coded keywords guide Lisa in considering various aspects of the dessert stand, such as functional, behavioral, and structural constraints. While refining her design, Lisa remembers the "polished surface" from the vinyl and decides she wants to incorporate this detail. Since ProductMeta has not suggested this property, she adds it as an attribute by clicking the "Plus" button (Fig. 7. c). After clicking the "Generate" button, ProductMeta presents four distinct design proposals in the Proposal Creation Panel (Fig. 7. d). Lisa clicks "More Options" button within the Proposal Card to view a detailed description (Fig. 7. e). In the Design Proposal Construction Module, Lisa refines the design by selecting keywords in the Constraints Panel and Attributes Panel. She maintains the sturdy central rod from the dessert stand and the cool metal tonearm from the vinyl (Fig. 7. a,b). After clicking "Generate," the system presents four design proposals in the Proposal Creation Panel (Fig. 7. d). Lisa can see detail design description by clicking "More" on the Proposal Card (Fig. 7. e). To further refine the design, Lisa adds the "polished surface" property, which was not suggested by ProductMeta, by clicking the "Plus" button (Fig. 7. c). She can also upload images of both the target and the source for attribute elicitation (Fig. 7. f). After revising the keywords in both panels, Lisa clicks "Generate" again, prompting the system to present four new proposals based on her modifications. Once Lisa revises the keywords selected in both the Constraint and Attributes Panels and clicks the "Generate" button again, the system presents four new proposals based on her modifications.

If Lisa is not satisfied with the design proposals, Lisa can revisit the Mapping Exploration Module or the Source Exploration Module to try different mappings or explore additional sources. This flexibility allows Lisa to iterate and refine her design ideas before finalizing her proposal.

6 EVALUATING PRODUCTMETA

To evaluate ProductMeta's support for novice designers in ideating and creating product metaphors, we conducted a within-subjects study with 16 novice designers, comparing ProductMeta to a ChatGPT interface powered by GPT-4o.

We selected GPT-4o for comparison as: 1) Existing AI tools lack comprehensive support for metaphorical product design tasks. and 2) ProductMeta is built on top of GPT-4o and thus leverages the same model capabilities, but it incorporates additional features tailored to support metaphorical design tasks. These features include a structured metaphor creation pipeline, interactive interface, and design principles integration, which are absent in ChatGPT's basic interface. Comparing ProductMeta to GPT-4o allows us to evaluate the effectiveness of these additional design features in enhancing the ideation process for novice designers. 3) While other AI tools like MidJourney and Stable Diffusion excel in image generation, they lack integration with other design functions and require more technical expertise.

To control variables, we standardized prompt guidelines for generating images during the GPT-4o task. This ensured participants had equal familiarity with the tools. During the user experiment, we first observed how each system supported the different phases of ideation. Next, we assessed whether ProductMeta enhanced participants' creative processes and improved their final design outcomes. Lastly, we evaluated the usability of both systems. Our hypotheses are as follows:

- **H1:** Users will benefit from the system in the 3-steps of the metaphor process— resources exploration (H1a), mappings exploration (H1b) and design proposal construction(H1c).
- **H2:** ProductMeta can enhance user experience in creative processes, including perceived exploration (H2a), worth of effort (H2b), engagement (H2c), expressiveness (H2d), tool transparency (H2e), and collaboration (H2f).
- **H3:** ProductMeta is more useful for novice designers.

ID	Major	Design Exp. (Yrs.)	Metaphor Design Exp. (Times)	Familiarity with Large Language Models?	Familiarity with image generation using large
P1	Information Design	3	1	Fairly familiar	Not at all
P2	Painting	0	1	Not at all	Not at all
P3	Journalism (Minor in Digital Media Art and Design)	2	1	Fairly familiar	Fairly familiar
P4	Art History and Theory	3	2	Moderately familiar	Moderately familiar
P5	Craft Art	3	2	Slightly familiar	Somewhat familiar
P6	Ceramic Art and Design	2	2	Somewhat familiar	Somewhat familiar
P7	Industrial Design	1	2	Fairly familiar	Slightly familiar
P8	Innovation Design Engineering	3	2	Moderately familiar	Slightly familiar
P9	Architecture Design	3	1	Somewhat familiar	Moderately familiar
P10	Information Design	3	1	Somewhat familiar	Somewhat familiar
P11	Journalism (Minor in Digital Media Art and Design)	3	1	Fairly familiar	Moderately familiar
P12	Environmental Engineering	3	3	Moderately familiar	Slightly familiar
P13	Product Design	3	1	Moderately familiar	Moderately familiar
P14	Product Design	3	1	Moderately familiar	Moderately familiar
P15	Visual Communication Design	3	1	Slightly familiar	Somewhat familiar
P16	Product Design	2	2	Slightly familiar	Moderately familiar

Table 1. Participant demographics, including participant ID, major, design experience (in years), prior experience with product metaphor design (number of times), and familiarity with large models.

6.1 Participants

We recruited 16 novice designers (see Table 1; 14 female, 2 male; age range 20-40) through online advertising. Participants were required to hold a degree in Design or Art and have less than four years of academic. Additionally, they needed

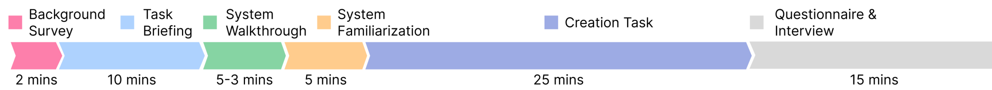

Fig. 8. The process of one task. Note that the first task involves a background survey and task briefing, which are excluded from the second task. Upon completing the second task, participants compare their experiences during an interview. The task concludes when participants submit their final design proposal in a Google Doc.

to have participated in at least one design project before. The study lasted about two hours, and participants were compensated with $20 for their time.

6.2 Tasks and Study Procedure

The overall process of the user study is shown in Figure 8. Participants were asked to perform design ideation tasks across two conditions: ProductMeta and GPT-4o. The task involved designing a metaphorical product, where they could choose two target products they found most interesting from the following list: 1) Kettle, 2) Toothpick holder, 3) Alarm clock, 4) Dessert stand, 5) Kitchen box, 6) Egg cup/holder, and 7) Cloth hanger. These items were selected from our dataset as representative of common metaphorical product designs within the home accessory category. To further motivate participants, an additional $10 compensation was offered to those with the highest performance. To avoid bias, the order of tools (ProductMeta and GPT-4o) and topics was counterbalanced, meaning participants alternated between starting with each tool and task. The user study was conducted via Zoom, with participants accessing ProductMeta through their web browsers and sharing their screens throughout. Throughout the metaphor creation sessions, they shared their screens continuously.

Before the experiment, participants were given a 10-minute briefing to familiarize themselves with the definition and process of product metaphor design. During the first five minutes of each round, the experimenter introduced the main features of the tool, guiding participants step by step through the process of creating a product metaphor. After that, participants had five minutes to explore the tool and become familiar with its functionality. Once this phase was complete, participants were presented with a list of potential targets. After selecting their target product, target users, and usage environment, participants began their task.

During the creation task, participants completed two metaphor product design tasks. They were asked to submit one design proposal after the experiment. Their final output included: 1) a product design description (must contain the names of the source and target, the connection between source and target, the mapping strategy, and the conveyed meaning or emotions), 2) a slogan, and 3) an image generated by the tool that best represented their design. The task took about 25 minutes and finished once participants submitted their final design proposal. All study sessions were audio and video recorded with the consent of the participants.

After each round, we conducted a 15-minute semi-structured interview to understand participants' creative process and how the tool influenced their ideation process. Following the final round, they were also asked to compare their experience in both systems.

Additionally, they reviewed their design processes to recall and compare them with their traditional design workflows without AI. The interview questions are in Appendix A.3.

6.3 Measures

After each round, participants completed a survey evaluating their satisfaction with the system across different stages of the ideation process: source inspiration (H1a), mapping exploration (H1b), and visual concept construction (H1c). They rated the system's support at each stage using a 7-point Likert scale. Additionally, we assessed participants' overall perception of the creative process using the Creativity Support Index (CSI) (H2) [23], and the tool's usability with the System Usability Scale (SUS) [7, 13].

To assess how ProductMeta supports exploration during metaphor creation, we manually coded screen recordings and analyzed the time distribution and interaction counts across three key phases: source exploration, mapping exploration, and visual construction. We also examined phase switches, iteration steps, and the number of concept proposals generated before finalizing visuals. Details of the coding rules and definitions are provided in Appendix.

We also evaluated the quality of the design outputs with four design experts. Each participant's final design proposal was selected for evaluation, resulting in a total of 32 designs (1 proposal × 16 participants × 2 conditions). The experts included two individuals with master's degrees (one in visual communication design and one in industrial design) and two with PhDs in industrial design, all possesssing over seven years of design experience. During the assessment, we presented the ProductMeta and GPT-4o design results side by side, and the experts were asked to evaluate and compare each design through a semi-structured interview. The interview was conducted via Zoom, with each session lasting approximately one hour. Experts were compensated with $50 for their participation.

6.4 Result

Results showed that ProductMeta addressed the key challenges of metaphorical design by: (1) enabling broader exploration of diverse ideas during source and mapping phases; (2) assisting in understanding and refining target constraints during proposal construction; and (3) facilitating a more thorough exploration and refinement of mapping styles during the proposal construction phase. Additionally, ProductMeta fostered exploration and enhanced the perceived value of effort in the creative process. Quantitative findings revealed that ProductMeta supported iterative exploration across all phases, whereas GPT-4o encouraged a more linear process focused on visual refinement.

6.5 Support for different Metaphor stages

To answer H1, we examined survey responses and interview results across 3 steps of metaphor design: (1) Source Finding; (2) Exploring property mappings and (3) doing visual concept constructions. Wilcoxon signed-rank tests were used to compare the differences in satisfaction with support from the two systems.

6.5.1 Inspire more Creative Source. As shown in Figure 9, ProductMeta was significantly more effective in fostering creative and diverse source exploration ($M = 5.81$, $SD = 0.88$) compared to the GPT-4o system ($M=5.12$, $SD=1.22$, $p=0.022$, $W=9.0$). When asked about the creativity and diversity of suggestions, ProductMeta ($M=6.13$, $SD=0.88$) also outperformed the GPT-4o ($M=5.6$, $SD=1.02$, $p=0.059$, $W=20.0$). Thus, H2a is accepted.

Many participants reflected, "The system inspires me with novel and diverse source suggestions, offering more context-specific recommendations than GPT-4o" (P1-4, P8, P10, P11, P13, P15, P16). Some participants mentioned that ProductMeta provides unexpected connections between the target and source: "Even if their relationship seems distant at first glance, the system provides reasonable connections and helps me think out of the box" (P3, P4, P13). On the other hand, a few participants argued that some sources could be seen as "novel but irrelevant or far-fetched" (P2, P12).

Also, some participants noted that ProductMeta suggests tailored sources based on design context and target users. P8 noted: "I wanted to design an alarm clock for the deaf, and this system suggested some non-audio interactions like simulating an earthquake or butterfly wings fluttering, which is suitable for my target users." In contrast, some participants reported that the GPT-4o's suggested metaphors were too abstract or lacked a strong connection to the target product. (P10, P12).

Compared to traditional design process, ProductMeta expands the scope of ideation by suggesting sources beyond what is typically accessible through traditional search engines (P1,P3,P11,P15).

Regarding the interaction design, nine participants appreciated the Metaphor Cards in source explorations. "The high-quality, relevant visuals effectively sparked my imagination and encouraged me to intuitively understand the connections between the source and the target." (P1, P4, P10, P13, P16).

6.5.2 Support Exploration of Mappings. As shown in Figure 9, participants found that ProductMeta was significantly more effective in helping them discover diverse mapping strategies (M=6.25, SD=0.83) compared to the GPT-4o system (M=4.88, SD=1.32; p=0.004, W=0.0). Additionally, participants perceived that ProductMeta better supported them in exploring multi-sensory mapping strategies (M=6.06, SD=0.63) compared to the GPT-4o system (M=5.5, SD=1.06; p=0.004, W=6.0). Thus, H1b is fully supported. Most of the participants mentioned the system guided them to think from multiple dimensions (12/16). "It provides several sensory dimensions that help me categorize and think systematically." (P1, P3, P6-8, P10, P15, P16). "Except for what it looks like and uses like, the system also provides what it sounds like or smells like, guiding me to explore dimensions I hadn't thought of." (P7, P8, P10, P16). P1 mentioned, "I initially considered only one sense, but the system suggested many more possibilities." The design of Property Pairing Panel is thought to be more intuitive and effective when exploring mapping strategies. P3 mentioned: "GPT often provides complex explanations with lengthy, unnecessary details. ProductMeta simply presents the connections and allows me to filter and manage them. " In addition, participants are satisfied with the function of uploading images for properties analysis, as it helps them gain deeper insights when analyzing products. P10 noted, "I wanted the egg cup to function like a volcano, which could heat up and visualize the volcanic eruption effect. So, I uploaded an image highlighting the heating mechanism of a volcano, and the system was able to recognize the features such as glowing lava and heat generation."

Category	Factor	ProductMeta		GPT-4o		Statistics			Hyposis
		Mean	SD	Mean	SD	Statistics	*p*	Sig,	
Creativity Support Index	Exploration	8.13	1.83	6.63	2.09	20.5	0.0216	*	*H2a* accepted
	Engagement	8.19	1.74	6.06	1.43	5.5	0.0119	*	*H2b* accepted
	Expressiveness	6.94	2.11	6.06	1.43	21	0.0418	*	*H2c* accepted
	Worth effort	7.19	2.19	6.38	1.62	21	0.0421	*	*H2d* accepted
	Tool transparency	6.69	2.08	5.56	1.80	11.5	0.0148	*	*H2* eaccepted
	Collaboration	7.06	1.78	6.69	2.08	41	0.3749	-	*H2f* rejected
Usability		73.00	11.13	67.50	12.10	18	0.0859	+	*H3* slightly accepted

Table 2. Survey results of self-perceived Creativity Support Index and Usability. ($-: p > .100, +: .050 < p < .100, *: p < .050, **: p < .010, ***: p < .001$)

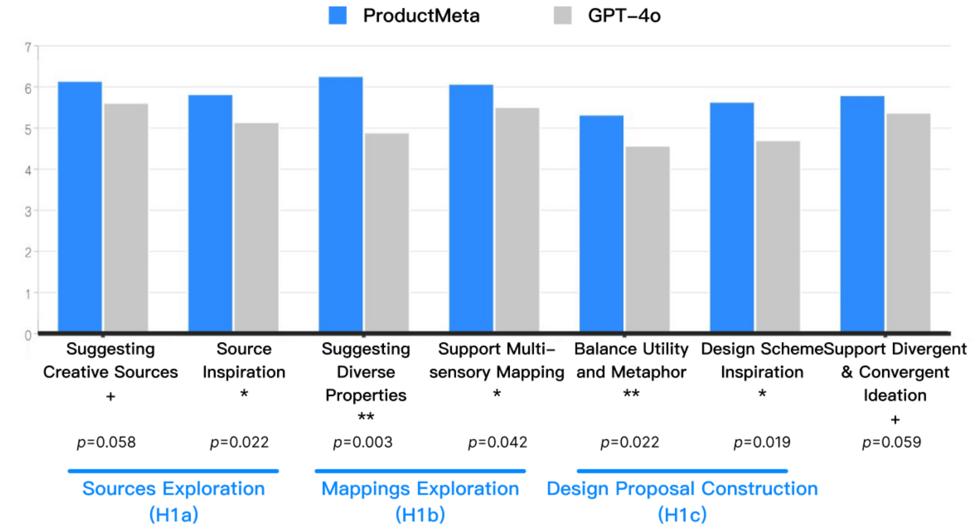

Fig. 9. Survey results on user-perceived satisfaction showed that ProductMeta significantly supports creative source discovery, exploration of diverse mapping strategies, and adjustment of product constraints to meet designers' needs. The findings are based on mean values.

6.5.3 Support design proposal construction. The investigation revealed that proposals provided by ProductMeta is significantly more inspiring (M=5.63, SD=1.05) than the GPT-4o system (M=4.69, SD=1.31; p=0.019, W=13.0). Additionally, ProductMeta better supports participants in balancing target constraints with metaphorical expression (M=5.31, SD=1.04) compared to the GPT-4o system (M=4.56, SD=1.32; p=0.022, W=1.0). Therefore, H1c is accepted.

According to the interview, some participants (6/16) mentioned that the Constraints Panel and the Source Attributes Panel support more flexible and systematic adjustments to the design proposals. "It can analyze my uploaded images and convert them into AI-understandable prompts, often extracting richer features than I had initially thought."(P4)." "The color-coded categories helped me explore design needs from multiple dimensions. It encouraged more systematic thinking and helped me focus on the concrete product image." (P9, P10, P11)

When evaluating the impact of outputs on participants' creativity, the design proposals generated by our system were considered more creative and inspiring (M=5.94, SD=1.34) compared to the GPT-4o (M=4.88, SD=1.89, p=0.023, W=10.5) (see Table 9). Some participants appreciated the diversity and inspiration provided by the proposals, noting that the four design options offered a range of directions (P3, P4, P5, P10). As P10 remarked, "It not only provides the rendering but also sketches and even a storyboard". Also, "This image inspired me to think from different angles, like transforming a modern candy cane into an artistic installation where the hanger resembles a sculpture in the images" (P4, P15).

However, regarding image generation, some participants mentioned that the images did not always align well with the keywords they selected, particularly when dealing with complex mappings - a limitation of current MLLMs. For example, P13 mentioned, "I aimed to incorporate light elements of the 'sunrise' into an alarm clock, such as a projection device that gradually brightens to simulate sunlight. However, the outcome did not fully align with my initial vision."

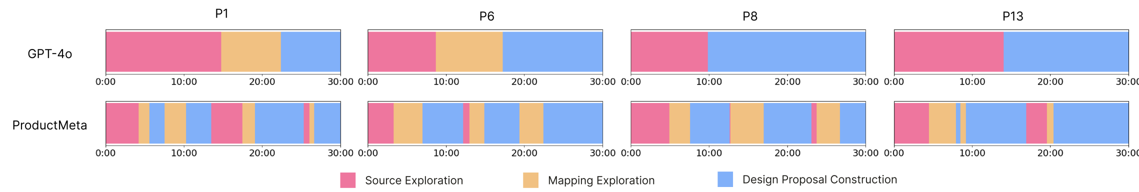

Fig. 10. Time distribution across three phases for P1, P6, P8, and P13, based on manually encoded screen recordings. Participants using ProductMeta showed more frequent phase switching, while P8 and P13 skipped the mapping exploration phase when using GPT-4o.

6.6 Impact on User's Creative Process

We assessed the system's ability to foster creativity using the Creativity Support Index (CSI) [23]. Participants rated their ideation experience with both ProductMeta and GPT-4o system on a 10-point Likert scale after each session. Wilcoxon signed-rank tests were conducted to determine if there was a statistically significant difference in perceived creativity support between ProductMeta and GPT-4o.

According to Table 2, participants prefer ProductMeta significantly more than GPT-4o regarding exploration (η^2 = .39), worth of effort (η^2 = .38), engagement (η^2 = .66), expressiveness (η^2 = .37), and tool transparency (η^2 = .54); H2a, H2b, H2c, H2d, and H2e are accepted, while H2f is rejected. Overall, H2 is partially accepted. It suggests that ProductMeta fosters users' experience in exploring various metaphor ideas, increases their sense of efficiency, and fosters a more immersive and engaging task environment.

6.6.1 Sense of Exploration. According to the interview, there is a distinction between ProductMeta and current MLLM tools. 14 participants viewed ProductMeta as a creativity tool, ideal for ideation, while GPT-4o was seen more as an efficiency tool, helping refine and implement existing ideas. Meanwhile, 6 participants emphasized that ProductMeta allowed for a freer exploration of different ideas and allow them to smoothly switch between divergent and convergent thinking (P3, P4, P7, P10, P11, P15).

Metrics	Phase Switches		Iteration Steps		Number of Concept Proposals	
	Mean	SD	Mean	SD	Mean	SD
GPT-4o	2.00	1.46	0.19	0.73	2.43	2.53
ProductMeta	6.43	3.18	2.57	1.92	7.21	3.36

Table 3. Comparison of key metrics between GPT-4o and ProductMeta: phase switches, iteration steps, and number of concept proposals. The definitions of the metrics are: (1) Phase switches: The number of times participants transitioned between the three phases. (2) Iteration steps: The step count during iterative backtracking, with one step indicating a return to the previous phase and two steps indicating a return to an earlier phase (e.g., from design proposal construction to source exploration), reflecting the extent of participants' iterative process. (3) Number of concept proposals: The quantity of mapping ideas generated by participants before finalizing their visual output. For further details, please refer to the appendix. Participants using ProductMeta demonstrated significantly higher values across all three metrics, suggesting greater engagement in exploratory and iterative workflows.

Supporting Non-linear Iteration. Participants using ProductMeta exhibited a non-linear and iterative workflow, with an average of 6.43 phase switches (SD=3.18) and 2.57 iteration steps (SD=1.92), frequently revisiting earlier phases

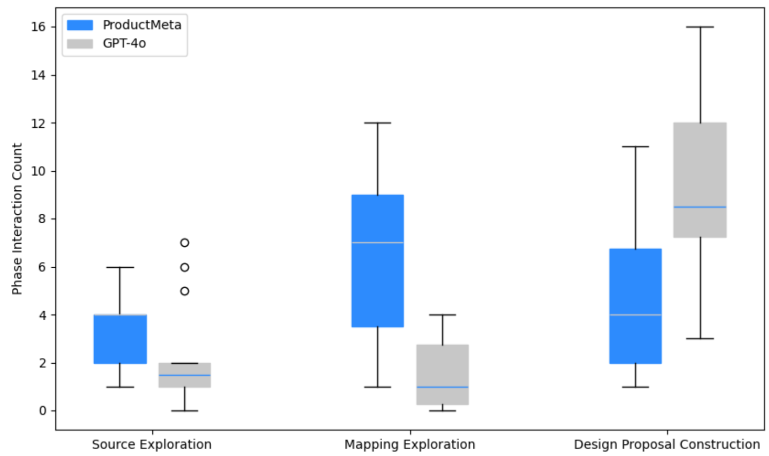

Fig. 11. Interaction counts across the three phases. Participants engaged more in early exploration phases (source and mapping) when using ProductMeta, with fewer interactions during the design proposal construction phase.

for ideation and refinement. In contrast, GPT-4o users followed a more linear process, with significantly fewer phase switches (mean=2.00, SD=1.46) and minimal iterations (mean=0.19, SD=0.73), apart from one instance (P2). P4 described this iterative approach in ProductMeta as a dynamic "diverge-converge-diverge" cycle, in contrast to GPT's more straightforward divergence-to-convergence process.

Avoiding Step-skipping. Another notable finding was the tendency of GPT-4o users to bypass early exploration phases. Some participants skipped critical steps such as mapping exploration, directly transitioning from source generation to visual construction (e.g., P8, P13 in Figure 10). This shortcutting hindered thorough idea generation and led to additional refinement in later stages. In contrast, ProductMeta encouraged users to engage comprehensively with all three phases. P3 noted that making the mapping process explicit allowed for “cross-fertilization of ideas", while P16 observed that this structure enabled the systematic consideration of diverse mapping strategies and sensory aspects, which might be overlooked in traditional workflows.

Enhancing Concept Ideation. The data also reflected these patterns in concept generation: With ProductMeta, users generated more concept proposals (mean = 7.21, SD = 3.36) compared to GPT-4o (mean = 2.43, SD = 2.53). According to P15, ProductMeta's design “continuously presents different options," fostering sustained creativity and exploration, whereas GPT “encouraged a narrower focus on its generated outputs”.

Guiding Early-phase Interactions. Interaction analysis revealed that ProductMeta effectively guides designers to focus more on the early exploration phases. As shown in Figure 11, participants engaged in more interactions during the source exploration (Mdn: GPT-4o: 1.5, ProductMeta: 4.0) and mapping exploration phases (Mdn: GPT-4o: 1.0, ProductMeta: 7.0) compared to GPT-4o. However, they engaged less during the visual construction phase when using ProductMeta (Mdn: GPT-4o: 8.5, ProductMeta: 4.0). For instance, P10 recorded 4, 11, and 5 interactions in each phase, generating six distinct solutions—147% more than the average of 2.43 concept proposals generated by users with GPT-4o. This focus on early phases would provide more clarity for visual construction and require fewer adjustments.

6.6.2 Worth effort.

Efficiency and Flexibility in Constructing Proposals. Participants found it easy and flexible to revise and refine design proposals using ProductMeta (P1, P2, P8, P10, P11). For instance, P8 noted, "ProductMeta's interface, with features like clicking and selection, improved my workflow and reduced redundant actions." In comparison, she felt "burdened by the natural language interaction with GPT-4o." P11 also observed, "Especially in metaphor design, which I'm less familiar with, the tool significantly improved my thought process and completion speed." This highlights how ProductMeta's structured framework-based interface facilitates iterative design and improves cognitive efficiency.

Compared to traditional methods like sketching or modeling, ProductMeta enhances efficiency by rapidly generating design variations (P1, P3, P6, P8, P10, P12, P13, P15) and consolidating tools into a single platform (P15). As P3 remarked, "I can quickly see different variations, helping me decide the direction without manually sketching each possibility." While this approach saves time, some participants felt it lacked the fine-tuning control offered by traditional hand-drawing methods (P3, P7, P10, P16).

Streamlining Brainstorming Efficiency. Compared to traditional design processes, ProductMeta accelerates the ideation process. Traditional methods often require extensive data gathering from multiple sources, which is time-consuming and can leave novice designers overwhelmed and directionless (P1, P3, P6, P10, P11, P12). In contrast, ProductMeta provides diverse source inspirations, reducing the effort for extensive research and enabling faster ideation. As P1 noted, it is "a better choice for quick proposals.

6.7 System Usability

To evaluate the usability of both systems, we adapted the standard System Usability Scale (SUS) questionnaire [7]. A Shapiro-Wilk test shows no significant departure from normality for both GPT-4o system ($W(16) = 0.97$, $p = 0.878$) and ProductMeta ($W(16) = 0.94$, $p = 0.576$). Given the normality of the data, we perform a paired sample t-test to compare the usability scores of ProductMeta and GPT-4o. The results indicate a marginally significant difference in usability scores between GPT-4o ($M = 67.50$, $SD = 11.13$) and ProductMeta ($M = 72.95$, $SD = 12.10$), $t(15) = 1.47$, $p = 0.086$. While the effect size ($\eta^2 = .18$) indicates a moderate practical difference, the results suggest that ProductMeta tends to have higher usability scores than GPT-4o at a 0.1 significance level.

7 DISCUSSION

We discuss the implications of **ProductMeta**, analyze design outputs, compare framework-based and natural language interactions, explore human agency and collaboration patterns, and outline future directions.

7.1 Showcase Analysis

Here we present design proposals from several participants and corresponding feedback from experts. Figure 12 presents four ideas from P2, P3, P10, P11. Idea (a) by P3, generated with ProductMeta. It received unanimous praise from all experts (E1-E4) for its subtle and refined combination of target and source metaphors. E2 commented, "It aligns well with the minimalist aesthetic and conveys a modern, serene feeling that fits the product's purpose. " In contrast, another

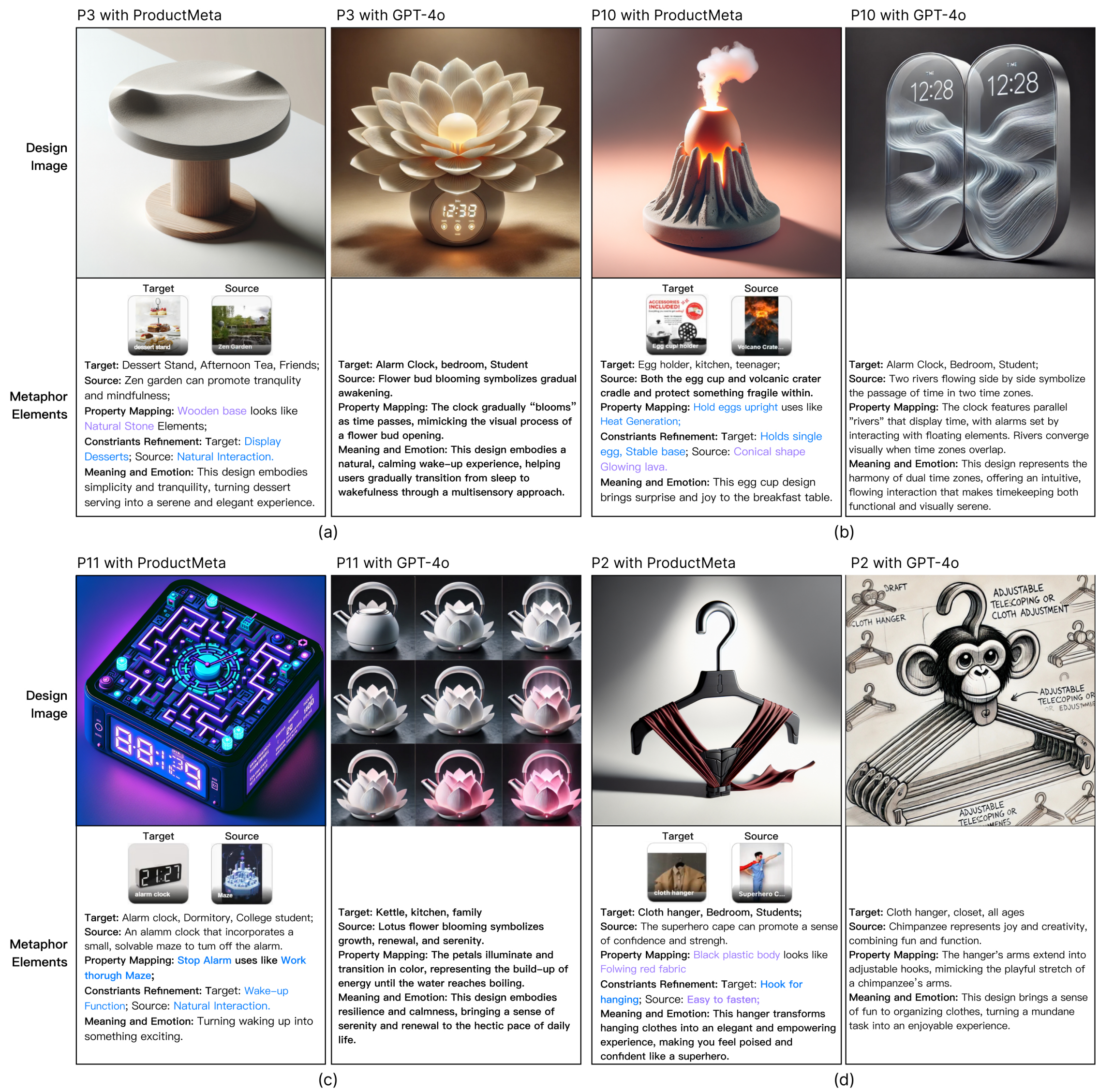

Fig. 12. Examples of 8 design proposals generated by 4 participants (P2, P3, P10, P11) using ProductMeta and GPT-4o. The target and source images in the "design elements" section were utilized by participants to develop their final concepts. For each participant's design, the left part is generated with ProductMeta, the right is generated with GPT-4o.

design created by P6 using GPT-4o simply pasted flower petals onto a clock, which was criticized for being "too abrupt and overly literal, leaving little room for imagination" (E1, E3).

Idea (d) by P2, also generated using ProductMeta, was appreciated for the source's narrative and contextual relevance to the target. E1 mentioned, "The idea of a suit hanger conveying a sense of heroism feels very fitting. By assigning the product a particular usage scenario, it becomes more engaging to me." Similarly, the idea (b) from P10 was recognized by experts as “It fits the context of a family breakfast, and children would love it.” (E1-E2). The idea (b) from P10 was praised for its more interactive mapping strategy, which resonated with expert E3: “The way you crack an egg mirrors a volcanic eruption, which aligns both in form and action.” This interactive metaphor was seen as successful in guiding user behavior through the design.

Idea (c) by P11 is an alarm clock design for college students. Initially, she selected a puzzle box concept but later shifted to a maze after experimenting with different mapping strategies. She explained, "The system recommended the maze, and I recognized a connection between the two." E3 commented: "The metaphor works because solving the maze engages your mind, making the wake-up process more than just a physical one." In contrast, the metaphor of a lotus and kettle in GPT-4o design was described by E3 as "overly generic" and the combination is not novel enough.

Overall, we find that ProductMeta suggests more concrete sources aligned with the target’s usage scenario and fosters a deeper connection with users. It also offers more subtle and flexible mapping strategies suited to each context.

7.2 Implications of Human Control and Agency

The integration of MLLMs has streamlined labor-intensive tasks in design workflows, such as research, drawing, and rendering, allowing designers to focus on higher-level decision-making. However, it has also raised concerns about diminishing human agency [44], suppressing creativity [89], and increasing homogeneity [22] in design outcomes. Our system prioritizes human control, ensuring AI supports not override designer decisions. As P5 stated, "I feel I have full control over the design process because, while AI lays the groundwork, I ultimately assess and refine the outputs.” In this section, we discuss the design choices in ProductMeta that promote human control and agency.

Structured Process. Firstly, we decompose the end-to-end process into 3 interactive modules, addressing the limitations of current MLLMs tools where users may struggle with understanding the underlying logic and influencing the branches. By navigating through these modules, designers can manage ideas more effectively and engage in cyclical divergent and convergent thinking—crucial for creative tasks. This design leads to a more iterative process shown in Table 3 and Figure 11.

Within the 3 modules, we offload the cognitive load encountered by designers to AI, allowing designers to focus solely on high-level decisions.

Framework-Based Design and Domain Knowledge in the Creative Process. During source exploration, when users lack initial inspiration, AI introduces creative sources informed by design principles (Norman’s three levels of emotional design) to encourage divergent and contextually relevant thinking. This aligns with research highlighting the value of external stimuli in early ideation [90]. Designers can also input their own ideas, maintaining control over exploration. Our user study shows that this interaction increases idea exploration and enhances the quality of generated sources (Section 6.6.1).

During mapping, designers take the role in selecting and combining ideas, while AI supports as an analyzer and refiner. The framework-based interface decomposes multi-sensory mapping perspectives into modular prompt templates, where AI analyzes and fills in the design elements (properties). This structure enables designers to experiment with multiple combinations and fosters a multi-sensory design thinking. As P10 noted, "ProductMeta’s framework helped me organize my thoughts and introduced a logical progression." Our study (see Table 3) shows that incorporating framework-based interaction during this phase leads to more idea exploration and a higher number of concept proposals.

During proposal construction, AI suggests key design considerations (balancing functionality with metaphorical expression). Designers can choose and configure these elements to generate proposals. This reduces the effort needed to craft prompts and mitigates the sunk-cost effect [88]. The framework-based interface also prompts designers to focus on key aspects of ideation rather than getting caught up in the details of image refinement.

Informed by design knowledge, our approach involves: 1) decomposing the creative process into phases, and 2) using a framework-based interface that enables AI to provide abstract elements rather than direct results. This approach supports informed human decision-making and enhances designer agency and control throughout the creative process [89].

7.3 Navigating AI-Generated Suggestions

Our study reveals the dual nature of AI-generated ideas. In the user study, participants occasionally encountered sources that appeared irrational or disconnected from the target product. However, these “outlandish” ideas also pushed participants to think beyond conventional boundaries. As P3 reflected, "I didn’t expect ProductMeta to suggest such a source. At first, I thought, ‘How could this work?’ But after exploring the idea, I realized it could be reasonable.”

On the other hand, over-reliance on AI without critical evaluation can result in worse design outcomes than those created solely by designers, leading to homogeneous, impractical, or misaligned designs [70, 71]. The core challenge is ensuring that designers evaluate AI-generated results, enabling them to leverage AI’s capability to generate unexpected ideas in more meaningful and impactful ways. While AI can spark creativity, the designer retains accountability to filter and refine these suggestions into meaningful solutions.

7.4 Generalization and Future Adaptations of ProductMeta

ProductMeta can be generalized to various design domains with revision of the specific domain principles. For example, in automotive design, designers often draw from sources (e.g., cheetahs) to convey speed. The existing Constraint Panel could be revised to support configurations such as drag coefficient and fuel efficiency. In architecture, mapping attributes could focus on spatial flow, lighting, material aesthetics, and user behavior. For more complex domains, future adaptations could involve balancing main metaphors and sub-metaphors, supporting both broad creative visions and fine-grained control [55].

While framework-based interfaces are useful, natural language tools excel in tailored interaction. As P13 noted, "GPT allows creative exploration and the freedom to ask what I want”, but this requires clarity of the design goal. Future work could explore a hybrid approach, combining structured frameworks with natural language interfaces, such as automatically adjust the Constraints and Attributes Panel when users request "render this design idea"—similar to systems like Stylette [56] and DynaVis [86].

To enhance user experience, future work should explore more intuitive interaction modes, such as integrating Sketches in the Loop for image refinement. Additionally, precision controls, like a slider to adjust the resemblance between source and target, could offer designers finer control. ProductMeta could also evolve into a plugin for tools like Figma, enabling seamless integration into the design workflow.

In our system, we decompose design tasks into prompt templates, enabling AI generate abstract design elements for the designer to select, thereby fostering an exploratory experience. This approach can also be applied to other design tasks like analyzing user profiles or forecasting new trends.

7.5 Limitation and Future Work

Our study has certain limitations that are worth mentioning. First, the user study was constrained by time, with participants spending only 25 minutes per condition to complete design tasks, which is significantly shorter than the actual design process. This limited duration prevents us from observing how design behaviors evolve over time. Second, the web response speed of the ProductMeta interface is significantly impacted by the model's generation speed, resulting in noticeable delays that affect the user experience. This latency reduces the fluidity of interactions and may cause users to lose focus and patience. Third, the AI-generated outputs in ProductMeta exhibit inherent biases originating from the training data. These biases reflect the pre-existing skewed patterns may inadvertently favor certain design aesthetics and concepts over others, thereby constraining the diversity of generated ideas. Finally, the system's metaphorical mappings are sometimes too broad and lack detailed mechanisms, which may hinder the novelty of the generated designs. For example, when combining a goose and an alarm clock, the system suggested integrating the sound of a goose, but a more innovative approach could involve a complex, physical mechanism, as seen in previous designs. This highlights the challenge of current MLLMs in understanding physical conditions and product functionality.

Therefore, future work could extend the user study duration and enhance the MLLMs' performance, especially in reducing latency, mitigating biases, and improving the generation of more sophisticated and contextually appropriate contents.

8 CONCLUSION

In conclusion, this research systematically explored the challenges faced by novice designers in metaphorical product design, identified gaps in existing creative support tools, and highlighted opportunities for improving the ideation and design process. To address these challenges, we introduced ProductMeta, an interactive creative support tool that integrates domain knowledge with a framework-based interaction design. ProductMeta facilitates the ideation, creation, and iteration of metaphorical product designs through human-AI collaboration, offering a flexible, non-linear process. The user study demonstrates its effectiveness in supporting creative process and 3 ideation phases, helping designers navigate complex mapping processes while maintaining control over the design outputs. Moving forward, we believe ProductMeta can inspire future research and development in the field of creative support tools, offering a more intuitive and effective design workflow for novice designers.

REFERENCES

[1] Amro Abbas, Kushal Tirumala, Dániel Simig, Surya Ganguli, and Ari S. Morcos. 2023. SemDeDup: Data-efficient learning at web-scale through semantic deduplication. arXiv:2303.09540 [cs.LG] https://arxiv.org/abs/2303.09540

[2] Mersha Aftab and Helen Agustin Rusli. 2017. Designing visceral, behavioural and reflective products. Chinese Journal of Mechanical Engineering 30 (2017), 1058–1068.

[3] Arjun R Akula, Brendan Driscoll, Pradyumna Narayana, Soravit Changpinyo, Zhiwei Jia, Suyash Damle, Garima Pruthi, Sugato Basu, Leonidas Guibas, William T Freeman, et al. 2023. Metaclue: Towards comprehensive visual metaphors research. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 23201–23211.

[4] Alessi. 2000. Anna G. https://us.alessi.com/products/anna-g-corkscrew?variant=33783678468236. Accessed: 2024-11-27.

[5] Shm Garanganao Almeda, JD Zamfirescu-Pereira, Kyu Won Kim, Pradeep Mani Rathnam, and Bjoern Hartmann. 2024. Prompting for Discovery: Flexible Sense-Making for AI Art-Making with Dreamsheets. In Proceedings of the CHI Conference on Human Factors in Computing Systems. 1–17.

[6] Wendy Anderson, Ellen Bramwell, and Rachael Hamilton. 2015. Mapping Metaphor with the Historical Thesaurus.

[7] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: adding an adjective rating scale. Journal of Usability Studies archive,Journal of Usability Studies archive (Apr 2009).

[8] Moshe Bar. 2004. Visual objects in context. Nature Reviews Neuroscience 5, 8 (2004), 617–629.

[9] LW Barsalou. 1999. Perceptual symbol systems. The Behavioral and brain sciences/Cambridge University Press (1999).

[10] Liliana Becerra. 2016. CMF design: the fundamental principles of colour, material and finish design. Frame Publishers.

[11] Alejandra Beghelli, Pablo Prieto, et al. 2020. Too many attributes!: Diminishing the cognitive load of metaphor generation in product design. In Proceedings of the Sixth International Conference on Design Creativity (ICDC 2020). 352–359.

[12] Alan F Blackwell. 2006. The reification of metaphor as a design tool. ACM Transactions on Computer-Human Interaction (TOCHI) 13, 4 (2006), 490–530.

[13] John Brooke. 2013. SUS: a retrospective. Journal of usability studies 8, 2 (2013).

[14] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, et al. 2023. Sparks of artificial general intelligence: Early experiments with gpt-4. arXiv preprint arXiv:2303.12712 (2023).

[15] Daniel Buschek, Lukas Mecke, Florian Lehmann, and Hai Dang. 2021. Nine potential pitfalls when designing human-ai co-creative systems. arXiv preprint arXiv:2104.00358 (2021).

[16] Hernan Casakin. 2011. Metaphorical Reasoning and Design Expertise: A Perspective for Design Education. Journal of learning design 4, 2 (2011), 29–38.

[17] Hernan Casakin et al. 2004. Metaphors in the design studio: Implications for education. In DS 33: Proceedings of E&PDE 2004, the 7th International Conference on Engineering and Product Design Education, Delft, the Netherlands, 02.-03.09. 2004. 265–273.

[18] Hernan P Casakin. 2006. Metaphors as an unconventional reflective approach in architectural design. The Design Journal 9, 1 (2006), 37–50.

[19] Hernan P Casakin. 2007. Factors of metaphors in design problem-solving: Implications for design creativity. International journal of design 1, 2 (2007), 21–33.

[20] 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-create visual metaphors. arXiv preprint arXiv:2305.14724 (2023).

[21] Hao Chen, Abdul Waheed, Xiang Li, Yidong Wang, Jindong Wang, Bhiksha Raj, and Marah I. Abdin. 2024. On the Diversity of Synthetic Data and its Impact on Training Large Language Models. arXiv:2410.15226 [cs.CL] https://arxiv.org/abs/2410.15226

[22] Liuqing Chen, Lingyun Sun, and Ji Han. 2023. A comparison study of human and machine-generated creativity. Journal of Computing and Information Science in Engineering 23, 5 (2023), 051012.

[23] Erin Cherry and Celine Latulipe. 2014. Quantifying the creativity support of digital tools through the creativity support index. ACM Transactions on Computer-Human Interaction (TOCHI) 21, 4 (2014), 1–25.

[24] Lydia B Chilton, Ecenaz 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. 1–14.

[25] 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. 1–14.

[26] Li-Yuan Chiou, Peng-Kai Hung, Rung-Huei Liang, and Chun-Teng Wang. 2023. Designing with AI: an exploration of co-ideation with image generators. In Proceedings of the 2023 ACM designing interactive systems conference. 1941–1954.

[27] 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. 1–25.

[28] WonJoon Chung. 2015. Theoretical structure of metaphors in emotional design. Procedia Manufacturing 3 (2015), 2231–2237.

[29] Nazli Cila. 2013. Metaphors we design by: The use of metaphors in product design. (2013).

[30] Richard Lee Davis, Thiemo Wambsganss, Wei Jiang, Kevin Gonyop Kim, Tanja Käser, and Pierre Dillenbourg. 2024. Fashioning Creative Expertise with Generative AI: Graphical Interfaces for GAN-Based Design Space Exploration Beter Support Ideation Than Text Prompts for Difusion Models. In Conference on Human Factors in Computing Systems-Proceedings.

[31] Pieter Desmet and Paul Hekkert. 2007. Framework of product experience. International journal of design 1, 1 (2007), 57–66.

[32] Gilles Fauconnier and Mark Turner. 1998. Conceptual integration networks. Cognitive science 22, 2 (1998), 133–187.

[33] Christiane Fellbaum. 2010. WordNet. In Theory and applications of ontology: computer applications. Springer, 231–243.

[34] Bruna Ferreira, Williamson Silva, Edson Oliveira, and Tayana Conte. 2015. Designing Personas with Empathy Map.. In SEKE, Vol. 152.

[35] Kenneth P Fishkin. 2004. A taxonomy for and analysis of tangible interfaces. Personal and Ubiquitous computing 8, 5 (2004), 347–358.

[36] Charles Forceville et al. 2008. Metaphor in pictures and multimodal representations. The Cambridge handbook of metaphor and thought (2008), 462–482.

[37] INTERACTION DESIGN FOUNDATION. [n. d.]. Split the Contents of a Website with the Pagination Design Pattern. https://www.interaction-design.org/literature/article/split-the-contents-of-a-website-with-the-pagination-design-pattern, year = 2000, note = Accessed: 2024-11-27.

[38] Giorgio Franceschelli and Mirco Musolesi. 2024. On the creativity of large language models. AI & SOCIETY (2024), 1–11.

[39] Dedre Gentner. 1983. Structure-mapping: A theoretical framework for analogy. Cognitive science 7, 2 (1983), 155–170.

[40] John S Gero. 1990. Design prototypes: a knowledge representation schema for design. AI magazine 11, 4 (1990), 26–26.

[41] Giorgio Giannone, Lyle Regenwetter, Akash Srivastava, Dan Gutfreund, and Faez Ahmed. 2023. Learning from invalid data: On constraint satisfaction in generative models. arXiv preprint arXiv:2306.15166 (2023).

[42] Hyoji Ha, Kwanghyuk Moon, Hyerim Joung, Hyegyeong Kim, and Kyungwon Lee. 2022. Exploration tool for effectively interpreting the visual metaphor process of sentiment visualization. In 2022 IEEE 12th Symposium on Large Data Analysis and Visualization (LDAV). IEEE, 1–2.

[43] Marc Hassenzahl. 2010. Experience design: Technology for all the right reasons. Morgan & Claypool Publishers.

[44] Jeffrey Heer. 2019. Agency plus automation: Designing artificial intelligence into interactive systems. Proceedings of the National Academy of Sciences 116, 6 (2019), 1844–1850.

[45] Paul Hekkert and Nazlı Cila. 2015. Handle with care! Why and how designers make use of product metaphors. Design Studies 40 (2015), 196–217.

[46] Jonathan Hey, Julie Linsey, Alice M Agogino, and Kristin L Wood. 2008. Analogies and metaphors in creative design. International Journal of Engineering Education 24, 2 (2008), 283.
[47] Jonathan HG Hey and Alice M Agogino. 2007. Metaphors in conceptual design. In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. 48043. 125–134.
[48] Jonathan Ho, Chitwan Saharia, William Chan, David J Fleet, Mohammad Norouzi, and Tim Salimans. 2022. Cascaded diffusion models for high fidelity image generation. Journal of Machine Learning Research 23, 47 (2022), 1–33.
[49] Jörn Hurtienne, Luciënne Blessing, et al. 2007. Design for Intuitive Use-Testing image schema theory for user interface design. In DS 42: Proceedings of ICED 2007, the 16th International Conference on Engineering Design, Paris, France, 28.-31.07. 2007. 829–830.
[50] Imoh M Ilevbare, David Probert, and Robert Phaal. 2013. A review of TRIZ, and its benefits and challenges in practice. Technovation 33, 2-3 (2013), 30–37.
[51] Mia Mohammad Imran, Preetha Chatterjee, and Kostadin Damevski. 2023. Shedding Light on Software Engineering-specific Metaphors and Idioms. arXiv:2312.10297 [cs.SE] https://arxiv.org/abs/2312.10297
[52] Phillip Isola and Ce Liu. 2013. Scene collaging: Analysis and synthesis of natural images with semantic layers. In Proceedings of the IEEE International Conference on Computer Vision. 3048–3055.
[53] Peiling Jiang, Jude Rayan, Steven P Dow, and Haijun Xia. 2023. Graphologue: Exploring large language model responses with interactive diagrams. In Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology. 1–20.
[54] 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. 1–15.
[55] Jeongyeon Kim, Sangho Suh, Lydia B Chilton, and Haijun Xia. 2023. Metaphorian: Leveraging Large Language Models to Support Extended Metaphor Creation for Science Writing. In Proceedings of the 2023 ACM Designing Interactive Systems Conference. 115–135.
[56] Tae Soo Kim, DaEun Choi, Yoonseo Choi, and Juho Kim. 2022. Stylette: Styling the web with natural language. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 1–17.
[57] Robert Kreuzbauer and Alan J Malter. 2005. Embodied cognition and new product design: Changing product form to influence brand categorization. Journal of Product Innovation Management 22, 2 (2005), 165–176.
[58] George Lakoff and Mark Johnson. 2008. Metaphors we live by. University of Chicago press.
[59] Ming-Huang Lin and Shih-Hung Cheng. 2014. Examining the "later wow" through operating a metaphorical product. International Journal of Design 8, 3 (2014).
[60] Dan Lockton, Devika Singh, Saloni Sabnis, Michelle Chou, Sarah Foley, and Alejandro Pantoja. 2019. New metaphors: A workshop method for generating ideas and reframing problems in design and beyond. In Proceedings of the 2019 Conference on Creativity and Cognition. 319–332.
[61] Nick Logler, Daisy Yoo, and Batya Friedman. 2018. Metaphor cards: A how-to-guide for making and using a generative metaphorical design toolkit. In Proceedings of the 2018 designing interactive systems conference. 1373–1386.
[62] Geke DS Ludden, Barry M Kudrowitz, Hendrik NJ Schifferstein, and Paul Hekkert. 2012. Surprise and humor in product design: Designing sensory metaphors in multiple modalities. Humor 25, 3 (2012), 285–309.
[63] Atefeh Mahdavi Goloujeh, Anne Sullivan, and Brian Magerko. 2024. Is It AI or Is It Me? Understanding Users' Prompt Journey with Text-to-Image Generative AI Tools. 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 183, 13 pages. https://doi.org/10.1145/3613904.3642861
[64] Thomas Markussen, Elif Özcan, and Nazli Cila. 2012. Beyond metaphor in product use and interaction. Design and semantics of form and movement (2012), 110.
[65] George A Miller. 1995. WordNet: a lexical database for English. Commun. ACM 38, 11 (1995), 39–41.
[66] Samar H Nawar, Marwa S Etawy, Gehad E Nassar, Nagham Mohammed, and Ahmed G Hassabo. 2024. The impact of cmf design on product design. Journal of Textiles, Coloration and Polymer Science 21, 2 (2024), 259–272.
[67] Kent KW Neo. 2010. Metaphor as conceptual tool in design. In 44th Annual Conference of the Architectural Science Association, ANZAScA 2010, Unitec Institute of Technology.
[68] Don Norman. 2007. Emotional design: Why we love (or hate) everyday things. Basic books.
[69] Andrew Ortony. 1979. Beyond literal similarity. Psychological review 86, 3 (1979).
[70] Vishakh Padmakumar and He He. 2023. Does Writing with Language Models Reduce Content Diversity? arXiv preprint arXiv:2309.05196 (2023).
[71] Samir Passi and Mihaela Vorvoreanu. 2022. Overreliance on AI literature review. Microsoft Research (2022).
[72] 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. 804–818.
[73] Savvas Petridis, Hijung Valentina Shin, and Lydia B Chilton. 2021. SymbolFinder: Brainstorming diverse symbols using local semantic networks. In The 34th Annual ACM Symposium on User Interface Software and Technology. 385–399.
[74] Lyle Regenwetter, Akash Srivastava, Dan Gutfreund, and Faez Ahmed. 2023. Beyond statistical similarity: Rethinking metrics for deep generative models in engineering design. Computer-Aided Design 165 (2023), 103609.
[75] Paul Ricoeur. 1973. Metaphor and the central problem of hermeneutics. Graduate Faculty Philosophy Journal 3, 1 (1973), 42–58.
[76] Dan Saffer. 2005. The role of metaphor in interaction design. Information Architecture Summit 6 (2005).

[77] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. 2022. Photorealistic text-to-image diffusion models with deep language understanding. Advances in neural information processing systems 35 (2022), 36479–36494.

[78] Donald A Schön. 1979. Generative metaphor: A perspective on problem-setting in social policy. Metaphor and thought 2 (1979), 137–163.

[79] Linda M Scott. 1994. Images in advertising: The need for a theory of visual rhetoric. Journal of consumer research 21, 2 (1994), 252–273.

[80] Jay A Seitz. 1997. The development of metaphoric understanding: Implications for a theory of creativity. Creativity Research Journal 10, 4 (1997), 347–353.

[81] Olivier Serrat and Olivier Serrat. 2017. The SCAMPER technique. Knowledge solutions: tools, methods, and approaches to drive organizational performance (2017), 311–314.

[82] Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In Proceedings of the AAAI conference on artificial intelligence, Vol. 31.

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

[84] Jakob Tholander and Martin Jonsson. 2023. Design ideation with AI-sketching, thinking and talking with generative machine learning models. In Proceedings of the 2023 ACM designing interactive systems conference. 1930–1940.

[85] Karl T Ulrich and Steven D Eppinger. 2016. Product design and development. McGraw-hill.

[86] Priyan Vaithilingam, Elena L Glassman, Jeevana Priya Inala, and Chenglong Wang. 2024. DynaVis: Dynamically Synthesized UI Widgets for Visualization Editing. In Proceedings of the CHI Conference on Human Factors in Computing Systems. 1–17.

[87] Thomas JL Van Rompay and Geke DS Ludden. 2015. Types of embodiment in design: The embodied foundations of meaning and affect in product design. International journal of design 9, 1 (2015), 1–11.

[88] Vimal Viswanathan and Julie Linsey. 2011. Design fixation in physical modeling: an investigation on the role of sunk cost. In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. 54860. 119–130.

[89] Samangi Wadinambiarachchi, Ryan M Kelly, Saumya Pareek, Qiushi Zhou, and Eduardo Velloso. 2024. The Effects of Generative AI on Design Fixation and Divergent Thinking. In Proceedings of the CHI Conference on Human Factors in Computing Systems. 1–18.

[90] Qian Wan, Siying Hu, Yu Zhang, Piaohong Wang, Bo Wen, and Zhicong Lu. 2024. " It Felt Like Having a Second Mind": Investigating Human-AI Co-creativity in Prewriting with Large Language Models. Proceedings of the ACM on Human-Computer Interaction 8, CSCW1 (2024), 1–26.

[91] HH Wang and Wan-Ju Liao. 2009. Applications of metaphor theory to product design. Proceedings of International Association of Societies of Design Research (2009).

[92] 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. 1–19.

[93] E Winner. 1982. Invented worlds: The psychology of the arts. Harvard University (1982).

[94] Shukang Yin, Chaoyou Fu, Sirui Zhao, Ke Li, Xing Sun, Tong Xu, and Enhong Chen. 2024. A survey on multimodal large language models. National Science Review (2024), nwae403.

[95] JD Zamfirescu-Pereira, Richmond Y Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny can't prompt: how non-AI experts try (and fail) to design LLM prompts. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 1–21.

A APPENDIX

A.1 Formative study

ID	Gender	Age	Major	Design Experience (Years)	Frequency of use of Large Models
P1	F	22	Product design	5	use as I work
P2	F	24	Product design	5	sometimes
P3	F	27	Graphic design	8	seldom
P4	F	23	Product design, Interaction design	5	use as I work
P5	M	23	Product design, Interaction design	5	daily
P6	F	20	Graphic design	2	no
P7	M	36	Product design	12	sometimes
P8	M	24	Product design	5	daily
P9	F	23	Product design, Interaction design	4	sometimes
P10	F	23	Graphic design	4	no

Table 4. Demographics of all the participants in the formative study. This table includes participants' ID, gender, age, major, design expertise, familiarity with large models, and frequency of large modal usage (ranked from 1 to 10).

A.2 System Performance Evaluation on Source Exploration Diversity and Cohesiveness

To systematically evaluate the technical performance of the source exploration module and its ability to generate diverse and cohesive sources for metaphorical product design, we conducted an experiment focusing on the module's output diversity and coherence.

We selected 20 common products from Amazon's Home & Kitchen category [3] as design targets. For each target, two methods were used to generate 100 potential sources: simple prompts to GPT-4o and ProductMeta. The sources were embedded into 1536-dimensional vectors using OpenAI's "text-embedding-3-small" model. To assess diversity, the embedded sources were clustered into 5 groups using k-means++, with the number of clusters chosen to suit the dataset size and ensure meaningful semantic differentiation. Our method is inspired by prior works on clustering for diversity and deduplication [1, 21]. Two diversity metrics were computed:

(1) *Inter-cluster distance*: the average distance between cluster centers, indicating the breadth of source coverage across distinct semantic categories.

(2) *Intra-cluster distance*: the average distance between points within a cluster, reflecting the semantic cohesiveness of sources in each category.

The results demonstrate that sources generated by ProductMeta achieved a significantly larger inter-cluster distance (mean: 0.513, SD: 0.088) compared to GPT-4o (mean: 0.452, SD: 0.048). At the same time, ProductMeta's sources exhibited a smaller intra-cluster distance (mean: 0.765, SD: 0.057) than GPT-4o's (mean: 0.797, SD: 0.027). These findings indicate that ProductMeta generates sources that are: (1) more diverse, spanning a broader range of conceptual categories; (2) more cohesive, maintaining strong within-category consistency.

[3]https://www.amazon.com/home-garden-kitchen-furniture-bedding/b/?ie=UTF8&node=1055398&ref_=nav_cs_home

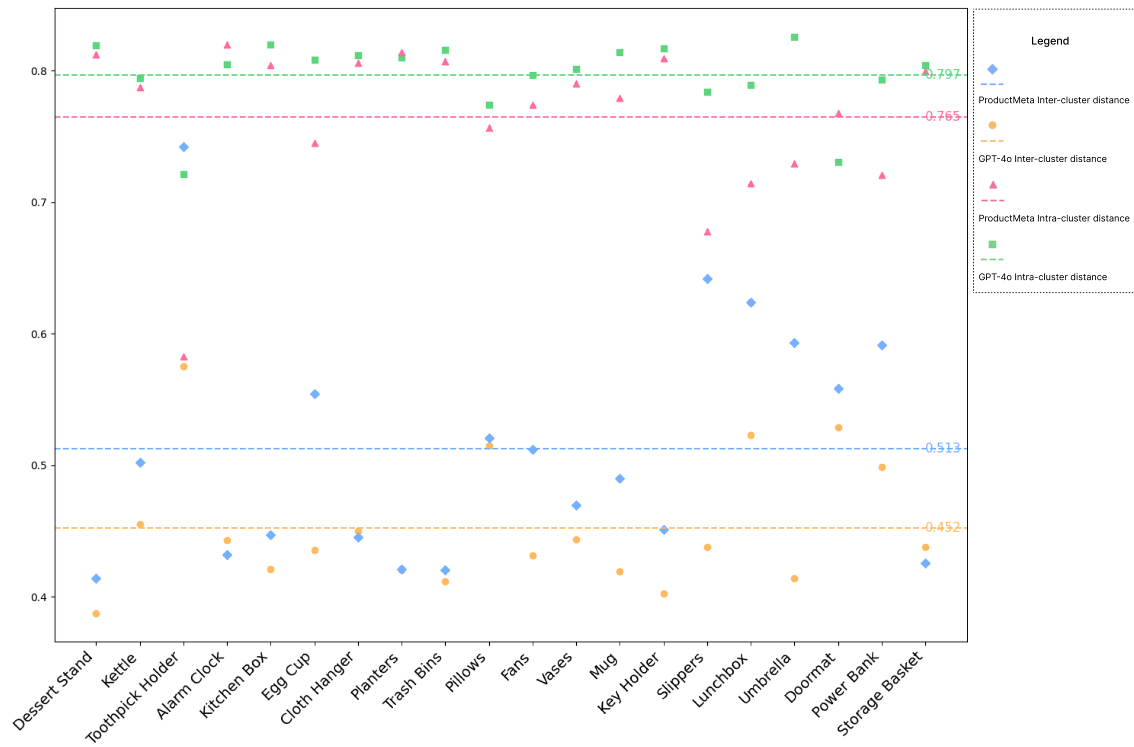

Fig. 13. Comparison of inter-cluster and intra-cluster distances for sources generated by GPT-4o and ProductMeta across 20 design targets. The inter-cluster distances reflect the diversity of sources, with ProductMeta achieving consistently higher values (mean: 0.513) compared to GPT-4o (mean: 0.452). The intra-cluster distances indicate semantic cohesiveness, with ProductMeta showing lower values (mean: 0.765) relative to GPT-4o (mean: 0.797). These results suggest that ProductMeta generates sources that are both more diverse and more cohesive compared to GPT-4o.

For metaphorical product design, these metrics have significant implications. A higher inter-cluster distance indicates that the sources cover a wider range of conceptual domains, which is critical for metaphors as they often rely on bridging distant conceptual spaces. Meanwhile, a lower intra-cluster distance ensures that each conceptual domain is internally consistent, making the suggestions more coherent and easier for designers to interpret. These findings align with the Salience Imbalance Theory, which suggests that effective metaphors require high object similarity and high category dissimilarity [69, 93].

A.3 User study

These are the questions used for the semi-structured interview after the 2 experiments sessions with GPT-4o and ProductMeta tools.

(1) Did the system help you in discovering sources? Please provide examples.
(2) Was the system effective in helping you combine the source and target?
(3) Did the system-generated images, descriptions, and slogans stimulate your creativity?

(4) How well did the system assist you in adjusting and managing the product's functional constraints while expressing metaphors?

(5) Did the system help you maintain flexibility between broad, divergent thinking and convergent thinking?

(6) What was your favorite feature of the tool? Why? Please provide examples.

(7) What aspects of the system would you suggest enhancing?

After the two sessions, participants will be questioned:

(1) Which of the two systems is more helpful in helping you find the source?

(2) Which of the two systems is more effective in helping you combine the source and the target creatively?

(3) Compared with the two systems, which system generated the images, descriptions and slogans that inspired your creativity more?

(4) Please evaluate and compare the two design outputs, which do you think is more creative? Why?

(5) Can you describe the tools or methods you typically use when completing metaphorical product design tasks? Please outline the specific steps in your design process, excluding AI tools.

(6) How would you compare the system used in the experiment with your traditional design process in terms of inspiration, efficiency, and overall experience? What are the reasons for your preference?

(7) The system in the video is divided into three phases: source discovery (finding metaphors), mapping explorations, and design visualization. Please describe the differences between the system and your traditional process in each of these phases, regarding how the 2 processes influence your creativity, and sense of exploration.

(8) In your opinion, what are the advantages and limitations of the system used in the experiment compared to traditional design processes?

A.4 Screen Coding Rules

We manually coded the screen recordings to analyze participants' behaviors during the three design phases. The coding includes: (1) time distribution, (2) interaction counts, (3) phase switches, (4) iteration steps, and (5) concept proposals generated before the final output. Below are the coding rules for each measure:

A.4.1 Time Distribution. **For ProductMeta:** The three phases—source exploration, mapping exploration, and design proposal construction—correspond to distinct interfaces (see Section 5.3 User Scenario). **For GPT-4o:**

- Source exploration: Participants explore sources.
- Mapping exploration: Participants integrate source and target concepts.
- Design proposal construction: Participants refine the generated image.

A.4.2 Interaction Counts. **For ProductMeta:**

- Clicking "Generate" to invoke MLLM analysis or generate content: 1 interaction.
- Uploading images or ideas: 1 interaction.
- Scrolling through descriptions: **not** counted.

For GPT-4o:

- Sending a prompt: 1 interaction.
- Regenerating outputs: 1 interaction.

A.4.3 *Phase Switches.* The number of transitions between the three phases: source exploration, mapping exploration, and design proposal construction.

A.4.4 *Iteration Steps.* The number of backtracking steps that measure the extent of participants' iterative processes:

- Moving from **design proposal construction** to **mapping exploration**: 1 step.
- Moving from **design proposal construction** to **source exploration**: 2 steps.
- Moving from **mapping exploration** to **source exploration**: 1 step.
- Linear progression from **source exploration** to **mapping exploration** to **design proposal construction**: 0 steps.

A.4.5 *Number of Concept Proposals.* The number of mapping ideas generated before finalizing the visual output. Due to time constraints, not all ideas were visualized, so this reflects participants' ideation capacity.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009

Category	Survey Question	GPT-4o Mean	GPT-4o SD	ProductMeta Mean	ProductMeta SD	Statistic value (W)	*p-value*	η2
H1a (DG1)	Did the system provide a wide variety of creative source ideas in different angles?	5.60	1.02	6.13	0.88	20	0.0588	0.35
H1a （DG1）	Did the system inspire you in finding various of sources in different angles?	5.13	1.22	5.81	0.88	9	0.0229 *	0.59
H1b （DG2）	Did the system help you quickly find and try out creative ways of mapping strategies?	4.88	1.32	6.25	0.83	0	0.0037 *	0.78
H1b （DG2）	Did the system help you explore multi-sensory mapping strategies of the source and the target?	5.50	1.00	6.06	0.83	6	0.0422 *	0.65
H1c （DG3）	Did the system help you control the output in your design scheme and make the scheme align with your expectation?	4.56	1.06	5.31	1.04	1	0.022 *	0.76
H1c （DG3）	Did the system provide diverse and creative design schemes, inspiring you to come up with more ideas?	4.69	1.31	5.63	1.05	13	0.0194 *	0.51
H1 （DG4）	Did this system support you in maintaining a broad perspective (divergent) while working on different design ideas (convergent) ?	5.36	0.89	5.79	0.56	5	0.0593	0.65
H2a (CSI) exploration	It was easy for me to explore many diferent optionsand ideas without a tedious, repetitive interaction.	6.63	2.09	8.13	1.83	20.5	0.0216 *	0.39
H2b (CSI) worth effort	What I was able to produce was worth the effort required to produce it.	6.38	1.62	7.19	2.19	21	0.0421 *	0.38
H2c (CSI) engagement	I was very absorbed/engaged in this activity - I enjoyed it and would do it again.	6.88	1.96	8.19	1.74	5.5	0.0119 *	0.66
H2d (CSI) expressive	I was able to be very expressive and creative while doingthe activity.	6.06	1.43	6.94	2.11	21	0.0418 *	0.38
H2e (CSI) tool transparency	While I was doing the activity, the system "disappeared," and I was able to concentrate on the activity.	5.56	1.80	6.69	2.08	11.5	0.0148 *	0.54
H2f (CSI) collaboration	I was able to work together with others easily while doing this activity	6.69	2.08	7.06	1.78	41	0.3749	0.13
H2 (CSI)	ALL	38.19	8.52	44.19	9.38	27.5	0.0167 *	0.28
H3 Usability	l think if next time I have this demand, I would like to use this system frequently.	N/A	N/A	N/A	N/A	N/A	N/A	N/A
H3 Usability	I found the system unnecessarily complex.	N/A	N/A	N/A	N/A	N/A	N/A	N/A
H3 Usability	l thought the system was easy to use.	N/A	N/A	N/A	N/A	N/A	N/A	N/A
H3 Usability	l think that l would need the support of a technical person to be able to use this system.	N/A	N/A	N/A	N/A	N/A	N/A	N/A
H3 Usability	l found the various functions in this system were well integrated.	N/A	N/A	N/A	N/A	N/A	N/A	N/A
H3 Usability	l thought there was too much inconsistency in this system.	N/A	N/A	N/A	N/A	N/A	N/A	N/A
H3 Usability	l would imagine that most people would learn to use this system very quickly.	N/A	N/A	N/A	N/A	N/A	N/A	N/A
H3 Usability	l found the system very cumbersome to use.	N/A	N/A	N/A	N/A	N/A	N/A	N/A
H3 Usability	l felt very confident using thesystem.	N/A	N/A	N/A	N/A	N/A	N/A	N/A
H3 Usability	l needed to leam a lot of things before l could get going with this system.	N/A	N/A	N/A	N/A	N/A	N/A	N/A
H3 Usability	ALL	67.50	11.13	72.95	12.10	18	0.1660	0.17

Table 5. The survey results of the user study (n=16), along with the questions, p-values, and mappings to the research questions. ($- : p > .100, + : .050 < p < .100, * : p < .050, ** : p < .010, *** : p < .001$)