QuerySwitch: Supporting the Design Process by Balancing Vagueness through Large Language Models

Myungjin Kim
Department of Artificial Intelligence
Hanyang University
Seoul, Republic of Korea
myungkim331@hanyang.ac.kr

Bogoan Kim
School of Information and Communication Engineering
Chungbuk National University
Cheongju, Republic of Korea
bogoan@cbnu.ac.kr

Kyungsik Han*
Department of Data Science
Hanyang University
Seoul, Republic of Korea
kyungsikhan@hanyang.ac.kr

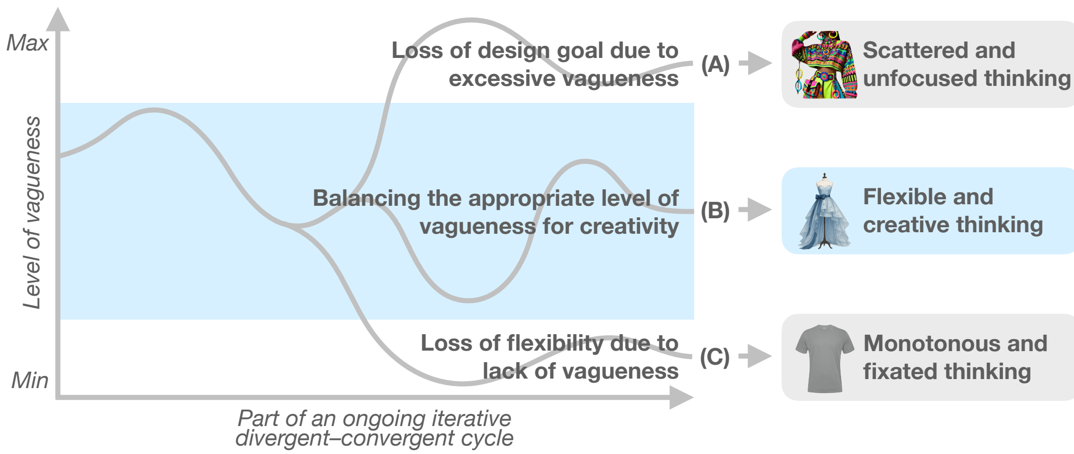

Figure 1: Conceptual diagram illustrating the relationship between creativity and vagueness during the iterative divergent–convergent process. The y-axis represents the level of vagueness, and the x-axis illustrates a segment of the ongoing iterative divergent–convergent cycle. When designers fall into excessive vagueness (A), their thinking becomes obscured and unfocused, drifting away from the core concept. Conversely, when vagueness is too low (C), thinking becomes fixated and monotonous, limiting opportunities for exploration. When designers maintain an appropriate level of vagueness (B), they are able to think flexibly and creatively, leading to satisfying design outcomes.

Abstract

Designers often regard vagueness as an essential aspect of creative work, as it fosters diverse interpretations and helps prevent fixation. Although large language models (LLMs) are increasingly viewed as a promising creative partner, designers struggle to productively incorporate vagueness into AI-supported workflows. To address this challenge, we present *QuerySwitch*, an interactive prototype that enables fashion designers to manage vagueness by flexibly switching between two distinct query-output modes. Findings from a user study show that *QuerySwitch* helps fashion designers balance vagueness, enhances the usability of LLMs in design tasks, and promotes creative exploration. This work contributes to HCI by (1) foregrounding a critical construct in human–AI collaboration, (2) demonstrating how interaction mechanisms can scaffold designer agency in LLMs use, and (3) articulating design principles—structuring exploration and preserving key query formulations—that extend to creativity-driven domains.

*Corresponding author

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3791729

CCS Concepts

• **Human-centered computing** → *User interface design*; **Empirical studies in HCI**.

Keywords

Human-centered AI, vagueness, creativity support tool, large language models, fashion design

ACM Reference Format:
Myungjin Kim, Bogoan Kim, and Kyungsik Han. 2026. QuerySwitch: Supporting the Design Process by Balancing Vagueness through Large Language Models. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3772318.3791729

1 Introduction

Throughout the creative process, designers often work with abstract mental representations such as rough sketches or generalized silhouettes of objects [10, 14, 68]. This vagueness does not hinder progress, rather provides a foundation for iterative divergent and convergent cycles. In the divergent phase, designers leverage vagueness to explore a wide range of interpretations and possible transformations. In the convergent phase, they maintain vagueness to synthesize and refine perspectives until the design becomes focused. Each phase may iterate multiple times, with the level of vagueness shifting accordingly. This balance helps generate diverse design strategies and prevents premature commitment to uncreative solutions [13, 69, 77]. However, when vagueness is poorly calibrated—whether excessive or insufficient—it can hinder the design process, leading to scattered, unfocused results or design fixation. Therefore, maintaining an optimal level of vagueness throughout the iterative divergent-convergent process is essential for effective progress and successful design outcomes [23, 24, 46, 53] (Figure 1).

Large language models (LLMs) have attracted increasing attention in design due to their capacity to generate tailored and novel creative outputs. Unlike earlier computational tools (e.g., Google Image Search and GANs), which were constrained by limited query input or fixed training datasets (e.g., narrowly defined keyword-based queries, pre-curated image corpora), LLMs offer far greater flexibility [28, 36, 39, 40, 73]. By enabling open-ended, context-rich textual prompts, LLMs can generate outputs that are novel and tailored to individual designers' goals [36]. This positions LLMs as a promising advancement for supporting creative design processes and offering adaptive, innovative results based on user input.

Building on this potential, many studies have examined how LLMs can support creative design tasks by generating visual outputs aligned with designers' contextual needs [3, 22, 41, 60, 72, 81]. Although these works demonstrate the value of context-aware image generation, they have primarily treated vagueness merely as a starting point, with limited attention to balancing it across the divergent–convergent process. However, a key and inherently subjective construct of the design process—in our work, vagueness—has not been integrated into the use of LLMs for human-AI collaboration. What remains missing is a research effort to bridge the exploratory nature of creative workflows with the prevailing interaction paradigms of LLM-based systems [32, 74, 76, 80]. This discrepancy underscores the need to identify system design requirements that can effectively support balanced vagueness during design tasks and to develop interaction strategies that provide designers with nuanced, context-sensitive assistance.

The goal of our research is to understand how balanced vagueness can be maintained in the fashion design domain during collaboration with LLMs. We first conducted a formative study with six senior fashion design professionals. The study yielded three key insights: (1) excessive vagueness caused by concept-deviating information during divergence; (2) vagueness minimization resulting from monotonous images during convergence; and (3) misalignment between the detailed input required by LLMs and the inherently vague inputs of the divergent-convergent process. Based on these insights, we established three design goals: (1) to prevent excessive vagueness by generating fashion design ideas anchored in the main concept; (2) to prevent the minimization of vagueness by producing diverse design outcomes from concrete inputs; and (3) to sustain different levels of vagueness during the divergent-convergent process through two forms of loosely defined queries. In response to these design goals, we developed *QuerySwitch*, an interactive system that allows fashion designers to dynamically switch between two distinct query modes—Abstract Query-Hierarchical Keyword and Parallel Query-Combination Image—depending on their current design context. With *QuerySwitch*, designers can actively choose the most appropriate query-output mode throughout the iterative divergent-convergent process, thereby enhancing creative exploration and ideation within design tasks.

We conducted a user study with 10 fashion professionals to evaluate the effectiveness of *QuerySwitch* compared to a baseline system. The results showed that *QuerySwitch* allowed participants to: maintain consistency while exploring varied ideas during divergence; support flexibility and prevent fixation during convergence; minimize query adjustment effort; and customize the system to their workflows. The results highlight five key contributions of LLM integration: (1) identifying structures of idea expansion from the core concept; (2) combining basic elements with concrete ideas; (3) preserving users' expressive styles in queries; (4) automating context recognition; and (5) generalizability to domains with fluid design elements.

In summary, this paper makes the following contributions:

- We identify the vagueness-oriented nature of the fashion design process, outline its associated challenges, and examine the potential and limitations of LLMs in addressing these challenges.
- We introduce *QuerySwitch*, an interactive prototype that leverages the capabilities of LLMs to switch between query and supports the iterative divergent–convergent process in design, allowing for balanced vagueness.
- We demonstrate that *QuerySwitch* allows users to effectively manage vagueness and improve the usability of LLMs. We further discuss the design implications for the development of interactive LLM-based creativity support systems.

Our study contributes to the growing body of research on AI-augmented creative support tools by demonstrating how the potential of LLMs can be harnessed to support creativity through the modulation of vagueness. Our approach, which grounds domain-specific characteristics (e.g., vagueness) in concrete design requirements, offers a potentially generalizable perspective for developing interactive systems that support creativity in other domains characterized by similar cognitive or design-oriented challenges.

2 Related Work

2.1 Balancing Vagueness in Creative Design Process

The uncertainty that occurs during the design process can be categorized into three types, based on the literature on design processes [1, 62, 67, 68] and linguistic classification [20, 26, 31]. The first type concerns ambiguity in communication, where a single expression may be interpreted in multiple ways even though only one meaning is intended, requiring shared understanding among designers or between designers and clients. The second type concerns uncertainty arising from a lack of knowledge, which occurs when designers lack the expertise or prior experience required to complete a task. In such cases, it becomes necessary to consult experts or seek out external knowledge sources. Finally, the third type is vagueness, a state in which the design has not yet been clearly defined, leaving specific elements indeterminate. Unlike the first two types, vagueness is deliberately leveraged by designers to foster exploration and creativity [67]. This paper focuses on vagueness as a key strategy within the design process.[1]

Vagueness is a recurring and integral feature throughout the divergent-convergent design cycle [14, 16, 67, 68]. In the early stages of ideation, designers often begin with vague and abstract mental images in which details remain undefined. This initial vagueness catalyzes divergent thinking, opening the space for multiple interpretations and transformations. During divergence, vagueness remains high, allowing fluid exploration across loosely structured concepts. In convergence, designers synthesize and refine ideas into more coherent directions, yet a degree of vagueness persists to keep multiple perspectives in play and avoid premature closure on uncreative solution. Only at the end of the convergence phase does the design become fully articulated and optimized. Thus, vagueness primarily characterizes the divergent-convergent process [16].

Operating within varying levels of vagueness is essential to creativity [13, 14, 69, 77]. However, imbalance can hinder progress. Excessive reduction of vagueness risks design fixation, narrowing exploration and producing uncreative outcomes. Conversely, excessive vagueness can obscure the main concept, increasing inefficiency and trial-and-error. Achieving the right balance enables designers to explore diverse ideas while maintaining coherence with the overarching design concept [15, 23, 46, 53]. When this balance is disrupted, either through premature specification or uncontrolled vagueness, both creativity and efficiency suffer.

In this paper, we define Balancing Vagueness as maintaining vagueness at an appropriate level during iterative divergent-convergent process, enabling openness to diverse exploration while preserving coherence with the main concept. There is a growing need for systems that help designers achieve this balance [8, 12, 15, 24]. Addressing this need, we present *QuerySwitch*, an interactive prototype designed to support the creative workflows by regulating vagueness across the design process.

2.2 LLMs in Creativity Support Tool

Computational technologies to support creativity in design have advanced considerably in recent years. In particular, the emergence of LLMs has enabled the generation of novel creative outputs by offering a more nuanced understanding of users' intentions and work contexts. Prior to the advent of generative models, image recommendation systems were available but limited in their effectiveness. These systems typically relied on one or two keywords, making it difficult to capture the subtle and complex design intentions [40, 61]. Advancements in GANs [17, 49], VAEs [30, 54], and Diffusion Models [7, 21, 47, 57] sought to overcome these limitations. The integration of CLIP [52] with diffusion models allowed more flexible and novel image generation based on natural language prompts. However, these models still struggled with accurately interpreting user intent [55], exhibited overfitting to training data, and often required careful prompt engineering [2, 36]. More recently, LLMs have further advanced this space [70]. DALL-E 3 leverages GPT-4 to improve image captioning and prompt formulation, producing higher quality images and attracting interest as a way to empower designers with richer exploration capabilities [2, 36].

Recognizing this potential, HCI researchers have begun investigating LLMs for creative exploration. However, their role in supporting balanced vagueness across the design process remains underexplored. Prior studies have largely treated vague and abstract ideas as merely the starting point of design. Some approaches refine vague queries into highly concrete prompts that generate fully specified images [3, 22, 60, 66, 72, 81], which can be useful for well-defined goals but risk suppressing productive vagueness. Others explore idea diversity (i.e., keywords and images) from vague inputs [5, 6, 41, 50, 65, 71, 83], but provide limited discussion of how systems can help users manage vagueness during divergence. Moreover, convergence is often framed as simply selecting one option among alternatives, overlooking the need for vagueness at this stage as well. As a result, few systems systematically support balanced vagueness across the entire divergent-convergent design cycle, despite its central role as a productive tool for creativity.

Although LLMs hold significant promise for enhancing creativity, their potential in modulating and sustaining vagueness remains unexplored. To address this gap, this study investigates a novel LLM interaction paradigm that supports balanced vagueness during the divergent-convergent design cycle.

3 Formative Study

Our focus was on the fashion domain, where managing vagueness throughout the divergent-convergent process is a key design strategy [11, 45, 56]. To identify considerations for human-LLM interaction design that support balanced vagueness and enhance the usability of LLMs in creative practices, we conducted a formative study with six fashion professionals, lasting around 60 minutes.

3.1 Participants and Recruitment

We recruited participants with comprehensive knowledge of the fashion design process from early ideation to final product development. To ensure this, we recruited fashion designers with at least one year of professional experience. This threshold was chosen because familiarity with both spring/summer and fall/winter collections—typically developed over a one-year cycle—is necessary for understanding the full scope of fashion design practices. Additionally, we required participants to have prior experience using generative tools such as ChatGPT and DALL-E to explore the potential and limitations of LLMs in the design process. A total of six participants (one male, five female), each with 2-4 years of professional experience, were recruited through an online community where fashion professionals regularly share fashion trends, job opportunities, and style tips. All procedures were approved by the

[1]Some studies use the terms vagueness, ambiguity, and uncertainty interchangeably. In this paper, we adopt a pragmatic approach by referencing previous literature that focuses on instances when designers lack a clear or concrete mental image.

Institutional Review Board (IRB) of the authors' institution. Participants provided informed consent and received $30 compensation.

3.2 Study Procedure

We followed the two key stages of the fashion design process [45]. The first stage involves creating a moodboard to establish the style direction for the subsequent sketch stage. The second stage focuses on refining the design through sketching, based on the visual concepts derived from the moodboard. We asked participants to share recent examples of moodboards and sketches they had created. Using these as reference points, we conducted semi-structured interviews regarding three central themes: (1) how initial design concepts evolve into more concrete forms, (2) the types of information participants typically seek, and (3) their prior experiences with generative tools such as ChatGPT and DALL-E, along with the challenges they encounter during the process.

Each participant then completed a 30-minute design task simulating real-world use of LLM-based tools. Participants were asked to create a mood board and a sketch. To initiate the design process from a state of vagueness, participants were asked to either imagine launching a new seasonal collection for their brand or reinterpret a previously developed concept with a new creative direction. Participants generated digital outputs with ChatGPT and DALL-E 3, created moodboards, and then produced final sketches using pencil and paper. During the task, they shared their screens with the researchers and were encouraged to think aloud, articulating their reasoning and providing real-time feedback on the tools. Afterwards, participants reflected on the potential and limitations of LLMs in fashion design.

All interview sessions were audio-recorded and transcribed. Two authors of this paper independently coded the transcripts and iteratively refined the coding results until reaching consensus on recurring themes. The scores for each category were higher than 0.76 (max = 0.93, avg = 0.87), indicating that the inter-coder reliability lies between "substantial" and "perfect" [34].

3.3 Formative Study Findings

Fashion designers progress through iterative cycles of convergence and divergence, moving from moodboards to sketches, while adjusting the level of vagueness to fit the design context. Designers are motivated to integrate LLMs into their workflow for two main reasons. First, novelty and originality are central to creative practice, and LLMs offer the ability to generate novel images that can inspire fresh directions. Second, given the diversity and complexity of design concepts, LLMs can produce tailored outputs even for highly specific topics (e.g., "pants with densely threaded ribbon-like textures"). Despite this potential, the use of LLMs in fashion design remains limited because of three challenges, detailed in Sections 3.3.1, 3.3.2, 3.3.3 (Table 1).

3.3.1 Divergent: Excessive Vagueness Caused by Concept-Deviating Information. Participants mentioned that during divergence, their inputs were typically abstract, which often led LLMs to generate outputs that deviated from the main concept and caused excessive vagueness. In general, designers expand ideas by holding the main concept in a rough mental form—sufficiently broad to embrace diverse directions but still anchored to prevent straying too far. This tendency applies to both the moodboard and sketch divergent phases. *"When I make moodboard, if the concept is feminine, I don't think about details. I just imagine a kind of feminine silhouette and explore various possibilities to develop it. Later, when exploring sketching ideas, it's the same. I only vaguely recall the moodboard's styling direction and broaden ideas without deciding on a specific design to sketch"* (P3).

Because designers' inputs remained abstract, LLMs frequently produced outputs that disrupted the designers' rough mental imagery of the concept. *"I wanted to design a sexy look, so I entered 'glamorous trendy,' but it gave me a tacky look with frills. I then switched to hip-hop and shifted again to a teenage style. Since the system kept giving me irrelevant outputs, I got confused about what kind of design I should even make"* (P1). Similarly, several participants (P2, P3, P4) reported that text keywords generated from abstract prompts were sometimes inspiring but also conceptually inconsistent. For example, a prompt "relaxed casual look." unexpectedly produced keywords related to a "bohemian look," These findings highlight that in the divergent stage, the inevitable reliance on abstract concepts can lead LLMs to generate information that deviates from the main concept, amplifying vagueness to an unproductive degree. These findings suggest the following design guideline:

- **DG1.** Prevent excessive vagueness by grounding fashion design ideas in the main concept.

3.3.2 Convergent: Vagueness Minimization Caused by Monotonous Images. Participants noted that during convergence, inputs related to concrete fashion elements became predominant, prompting LLMs to generate literal images and minimize vagueness. In general, as designers concretize concepts, they regulate vagueness by envisioning a mental image of concrete design element ideas, while leaving unspecified how these will be implemented in the outcome. Because it is somewhat concrete, this helps maintain consistency with the concept; and since the outcome is not yet fixed, it prevents fixation on a single solution and allows flexible consideration of diverse ideas. This is considered in both the moodboard and sketch convergent phases. *"Even if I make a moodboard with a pink, ruffle, dress style, I haven't yet decided which shade of pink to use or whether to put ruffles on the top or bottom. [...] Also, when sketching a pink ruffled blouse, I still don't know whether pastel pink or vivid pink would be better, so I have to consider different decorative aspects"* (P4).

Therefore, designers included concrete design elements in their queries; however, since they are concrete, LLMs are likely to generate monotonous and literal images. As a result, their concrete yet undefined mental imagery became fixed into a single idea. *"I wanted to get a variety of ideas for a hoodie and pleated denim skirt, so I asked for 'hoodie and short pleated denim skirt designs,' but instead of giving me diverse ideas, it only showed a single look—a gray hoodie with a short denim skirt. Even after several modifications, it kept showing something similar, so I ended up finalizing my design as a gray hoodie with a denim skirt"* (P5). This illustrates that during convergence, the unavoidable input of fashion elements leads LLMs to generate monotonous designs, thereby minimizing vagueness. These findings suggest the following design guideline:

- **DG2.** Prevent the minimization of vagueness by generating diverse design outcomes from concrete inputs.

Table 1: Summary of challenges, design goals, and system implementations identified through the formative studies on balancing vagueness with LLM tools. Each challenge corresponds to Sections 3.3.1, 3.3.2, 3.3.3.

Challenges	Design Goal	System Implementation
Excessive vagueness during divergence, as LLMs generate information that deviates from the main concept due to inevitable abstract inputs.	Prevent excessive vagueness by generating fashion design ideas based on the main concept.	Hierarchical keywords that contain sub-styles and categories derived from the main concept.
Minimal vagueness during convergence, as LLMs generate monotonous and overly literal images due to inevitable concrete idea inputs.	Prevent minimization of vagueness by generating diverse design outcomes based on concrete inputs.	Combinational images that contain diverse syntheses of user inputs and fashion components.
Misalignment between the level of detailed input required by LLMs and the levels of vague input inherent in the divergent-convergent process.	Support different levels of vagueness across divergent-convergent process by enabling two types of loosely defined queries.	Input abstract concepts as queries during divergence, and list fashion elements in parallel as queries during convergence.

3.3.3 Misalignment between LLMs' Detailed Input Requirements and Vague Inputs in the Divergent–Convergent Process. While all participants acknowledged the personalization capabilities of LLMs, they noted that crafting specific queries poses a significant challenge, particularly given the inherent vagueness. During divergence, abstract keywords such as "luxury minimal" were considered sufficient and even preferable, because they preserve openness to multiple interpretations. In contrast, during convergence, designers begin to envision specific ideas while the final outcome remains undetermined. This leads to relatively more concrete queries, such as the listing of fashion elements (e.g., "front chest slit, side slit, back cut-out detail"). Although these lists may appear specific, participants noted that the lists still keep a certain degree of vagueness, since the lists do not specify how the elements should be combined, weighted, or prioritized.

Participants emphasized that composing detailed prompts (e.g., "a white crop top layered over a white shirt, paired with a knee-length black leather skirt and white running-style sneakers.") rather disrupts the productive vagueness essential for creativity but also diminishes the perceived value of using LLMs. *"I once generated specific landscape images that were exactly what I needed for my personal portfolio [...] But in fashion design, if I can describe an outfit that specifically, it means I already have a complete vision in mind, and in that case, I don't need help from the system. I can just draw it myself"* (P2). These findings suggest the following design guideline:

- **DG3.** Support different levels of vagueness across divergent-convergent process by enabling loosely defined queries.

4 QuerySwitch

To address the identified design goals, we developed *QuerySwitch*, an LLM-based interactive system that helps designers balance vagueness throughout the divergent-convergent process.

4.1 QuerySwitch Main Functions

To preserve the vagueness characteristics of the divergent phase, users are encouraged to input abstract queries that reflect loosely defined design concepts, thus maintaining conceptual openness (DG3). In response, the system generates hierarchical keywords, expanding on the initial concepts to help designers explore diverse ideas while remaining anchored to the main theme (DG1). In the subsequent convergent phase, users construct parallel queries by combining hierarchical keywords, allowing them to retain the necessary degree of vagueness for convergence (DG3). The system then generates combinations of images that synthesize user design ideas into diverse potential outcomes. This supports convergence without premature fixation (DG2).

This structured interaction framework, implemented in both the moodboard and sketch stages, allows designers to iteratively explore and refine ideas while maintaining an appropriate balance of vagueness throughout the iterative divergent-convergent process (Figure 2). Designers can flexibly toggle between stages as needed, ensuring that interactions remain aligned with the iterative cycle.

4.2 QuerySwitch Interface

QuerySwitch features two mode selection buttons on the left: *moodboard* and *sketch*. When the moodboard button is selected, the interface provides input through an abstract text query and a parallel query. When the sketch button is selected, the input method shifts to an abstract image query and two types of parallel queries. Users can choose the query format most appropriate for their design phase by toggling the moodboard or sketch modes. All generated results remain accessible via scrolling and a history panel on the right side enables users to quickly revisit previously generated images.

4.2.1 Moodboard Stage. In the early moodboard stage, designers typically begin with abstract and loosely defined concepts. To accommodate this stage of high vagueness, *QuerySwitch* allows users to input queries in the form of one or two abstract keywords, reducing the cognitive burden of prompt formulation (Figure 3-A). Next, the system generates *hierarchical keywords* based on the user's query (Figure 3-B). Hierarchical keywords represent sub-styles related to the user's input, along with their associated fashion elements. For example, if the user enters a keyword "feminine," the system returns five sub-style interpretations—vintage feminine, chic feminine, romantic feminine, elegant feminine, and athleisure

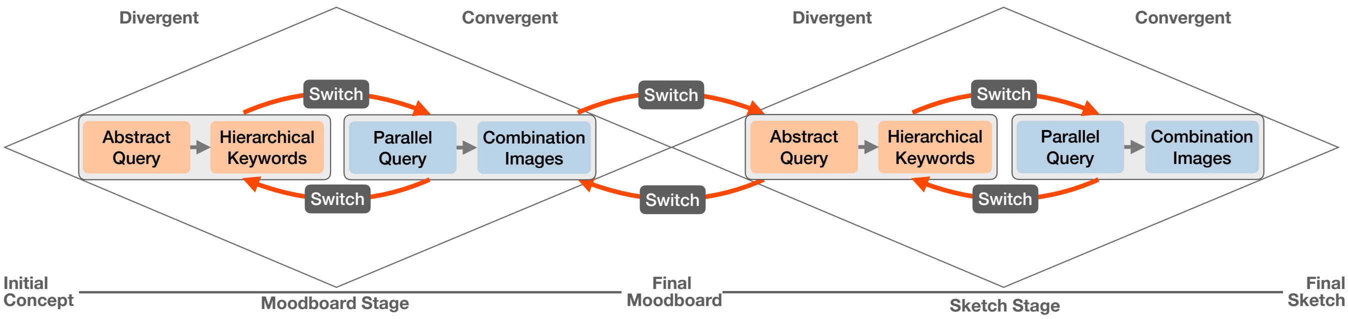

Figure 2: Process structure of *QuerySwitch*. *QuerySwitch* consists of two main modes—Abstract Query-Hierarchical Keywords and Parallel Query-Combination Images. Each mode is applied at both the moodboard and sketch stages. Balanced vagueness is managed throughout the iterative divergent-convergent process by switching between the modes. Design elements that support the divergent phase are highlighted in orange, while those that support the convergent phase are highlighted in blue.

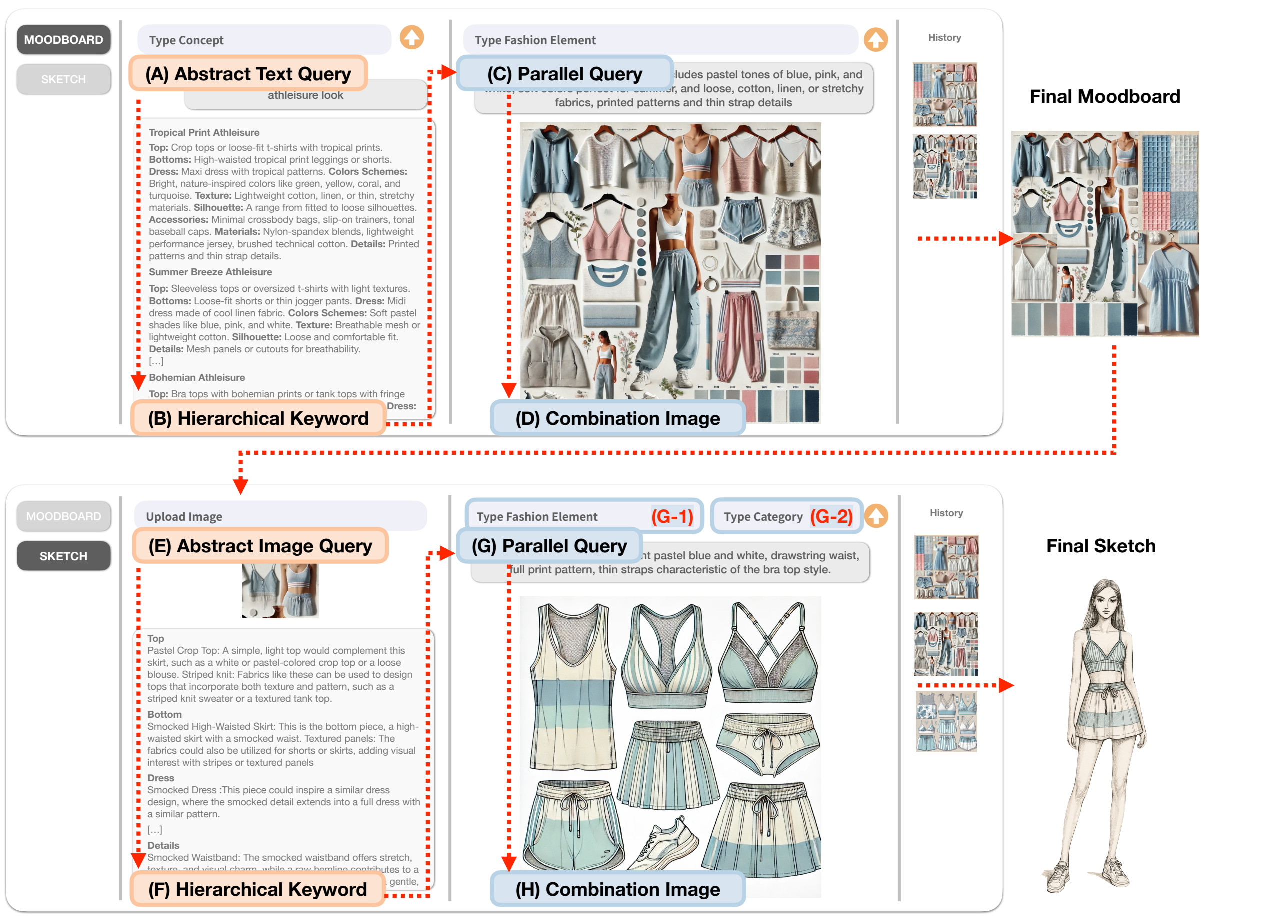

Figure 3: Illustration of the two stages of the design process: the "Moodboard Stage" (top) and the "Sketch Stage" (bottom). In the Moodboard Stage, (A) participants begin by entering an abstract design concept; (B) the system presents hierarchical results based on the input; (C) participants then provide parallel fashion elements; and (D) view the resulting moodboard combinations. In the Sketch Stage, (E) participants upload a moodboard image; (F) the system generates corresponding hierarchical results; (G) participants refine the prompt; and (H) the system presents fashion design combinations.

feminine. Each sub-style is further expanded across nine fashion categories—tops, bottoms, dresses, color schemes, textures, silhouettes, accessories, materials, and details—derived from established groupings commonly used in fashion literature and practice [29, 33]. Concrete fashion ideas are generated within each of these categories. This allows designers to explore a wide range of ideas while maintaining consistency with the main concept. In this way, designers can preserve the initial rough mental imagery associated with the concept and achieve a balance of vagueness.

Afterwards, designers formulate a follow-up query by selecting from the hierarchical keywords. To reduce the burden of writing detailed prompts, the system allows users to specify fashion elements in a loosely structured parallel query format (Figure 3-C). For example, a user might input "full skirts and tweed" from the vintage feminine sub-style, and "crop top and pink" from the romantic feminine sub-style. Based on these inputs, the system generates image combinations of moodboard components (e.g., color palettes, fabric samples, patterns) alongside user selection (Figure 3-D). The query "pink, tweed" might produce a color palette with toned-down pink, vivid pink, and tweed pink tops. This approach helps designers maintain a balanced level of vagueness during convergence by showing multiple ways to incorporate detailed inputs to the moodboard, which also helps preserve flexibility in the final outcome.

4.2.2 Sketch Stage. The overall style direction is established during the moodboard stage, and the sketch stage requires designers to further develop and concretize their ideas. The moodboard serves as a new abstract reference point, and exploration begins again with a stronger directional foundation. *QuerySwitch* allows users to upload the previously generated moodboard (Figure 3-E), reducing cognitive effort by accommodating the inherent vagueness of early sketching. Based on the uploaded moodboard, the system generates fashion element ideas organized into the same nine categories used in the moodboard stage—tops, bottoms, dresses, color schemes, textures, silhouettes, accessories, materials, and details (Figure 3-F). For example, the system might suggest "Dresses: Floral maxi dresses, short dresses with a dash of detail," or "Bottoms: Soft cottons and tough denims." This enables meaningful variation while keeping designers anchored to the moodboard concept.

As in the moodboard stage, designers then formulate a follow-up query by selecting from the suggested hierarchical keywords. To simplify prompt formulation, the system supports parallel input of fashion elements and categories (Figure 3-G). In particular, two input formats are available. The first accepts only fashion ideas; for example, entering "floral pattern, denim, cotton, pearls" into G-1 might receive images such as a denim floral dress or a denim cotton top. The second format allows users to pair ideas with specific categories. For example, when a user inputs "denim, pearls" in (G-1) and "pants" in (G-2), the system generates images of denim-and-pearl pants (Figure 3-H). This flexibility preserves vagueness during convergent sketching by showing multiple ways to translate detailed inputs into final designs, preventing premature fixation.

4.3 User Scenario

Scenario-based design is widely recognized for their effectiveness in articulating system use from multiple perspectives [4]. To illustrate how the existing tool (e.g., ChatGPT/DALL-E) and *QuerySwitch* apply to the practical challenge of balancing vagueness, we introduce Jenny, a fictional fashion designer. Figure 4 presents two scenarios: one in which Jenny struggles to balance vagueness using ChatGPT/DALL-E, and another in which she successfully manages it with *QuerySwitch*.

4.4 Technical Descriptions

QuerySwitch generates hierarchical keywords on the moodboard interface using OpenAI's GPT-4-turbo model, chosen for its strong performance in generating semantically relevant subcategories from user queries. The same model is employed in the sketch interface to generate hierarchical keywords from uploaded images, leveraging its multimodal capabilities to interpret visual input with high precision. For image generation in both the moodboard and sketch interfaces, we used the DALL-E 3 API, which efficiently produces high-quality images from structured prompts. To ensure consistency and conceptual coherence throughout the design process, the user's initial concept query is reused across both the moodboard and sketch stages.

During the frontend–backend interaction, all user inputs are transmitted to the backend in JSON format via a REST API and are inserted into predefined prompt templates. In the Moodboard interface, both the user's abstract text query and parallel query are directly placed into the prompt. In the Sketch interface, when a user uploads a moodboard image as an abstract image query, the image is stored in AWS S3, and its link is sent to the backend in JSON format and inserted into the prompt. For parallel queries in the Sketch interface, the system employs two prompt types: one for queries entered only in G-1, and another for queries entered in both G-1 and G-2. The system selects the appropriate prompt based on the user's input configuration. Detailed prompt structures for each case are provided in the Appendix.

5 Study Design

To evaluate the effectiveness of *QuerySwitch*, we conducted a 90-minute within-subjects user study with 10 fashion professionals. The study consisted of two rounds of moodboard creation and sketch tasks, conducted in two different system environments. Specifically, we aimed to answer the following research questions:

- **RQ1.** How do hierarchical keywords and combinational images support fashion designers in balancing vagueness during the creative process?
- **RQ2.** How does LLM-driven context integration facilitate diverse design exploration and workflow efficiency across the moodboard-to-sketch stages?

To address these questions, we employed a mixed-methods approach. Quantitative data were collected through three standardized instruments: the USE questionnaire, self-perceived experience, and the NASA Task Load Index (NASA-TLX). In addition, we quantitatively evaluated the consistency and complexity of users' input queries and the generated images. For qualitative insights, we captured system interaction logs and conducted semi-structured interviews to understand user behavior and perceptions.

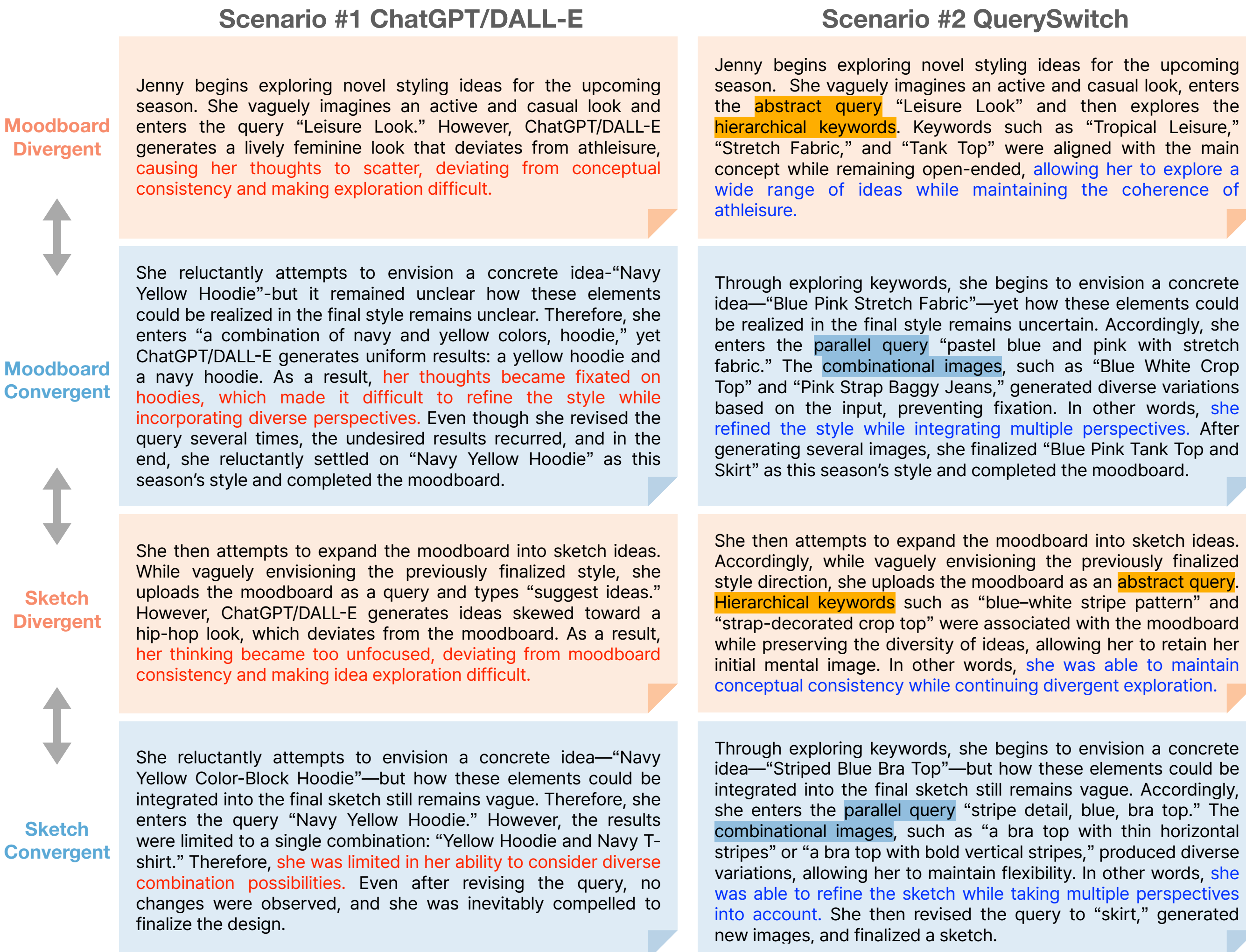

Figure 4: Comparative user scenarios of ChatGPT/DALL-E and *QuerySwitch* in fashion design ideation. Red highlights indicate the instance of vagueness imbalance observed with ChatGPT/DALL-E, whereas blue highlights indicate effective management of vagueness with *QuerySwitch* across divergent-convergent process.

5.1 Participants

We recruited professional fashion designers through an online community, following the same recruitment community as used in the formative study. All participants in this phase were newly recruited and had not participated in the earlier formative study. The inclusion criteria required at least one year of professional experience in fashion design to ensure familiarity with the full design process. A total of ten participants (four males, six females) were recruited, with between 1 and 5 years of industry experience. All participants reported prior experience using LLM-based generative tools, including ChatGPT, DALL-E 3, or Midjourney. The protocol was approved by the authors' institutional IRB, and all participants provided informed consent prior to participation. After the study, each participant was compensated $50 for their time.

5.2 Baseline

Our study does not aim to benchmark generative model performance. Instead, the focus is on evaluating how the design elements of *QuerySwitch* support the balancing of vagueness within the fashion design process. For the baseline, we chose ChatGPT with built-in DALL-E 3 functionality, as it employs the same underlying models (GPT-4-turbo and DALL-E 3) as *QuerySwitch*. This ensured consistent generative quality across both systems, allowing us to isolate and evaluate the effectiveness of *QuerySwitch*. In the baseline session, participants began with a vague prompt and iteratively refined it to generate both moodboards and sketches, receiving textual and visual responses in return. This setup mirrors real-world use of generative tools, enabling a realistic and meaningful comparison of how each system supports designers in managing vagueness during creative tasks.

Table 2: Details of the main study participants. Freelancer designer refers to participants who run their own personal brand, while brand designer refers to participants who are affiliated with a company. "Given Style (Q)" and "Given Style (B)" indicate the fashion styles assigned to each participant under the *QuerySwitch* and Baseline conditions, respectively.

ID	Job Description	Gender	Experience	Given Style (Q)	Given Style (B)
P1	Freelancer Designer	Male	2 years	Chic	Vintage
P2	Brand Designer	Female	3 years	Sporty	Feminine
P3	Brand Designer	Male	3 years	Sporty	Vintage
P4	Brand Designer	Female	2 years	Vintage	Feminine
P5	Freelancer Designer	Female	1 years	Vintage	Chic
P6	Brand Designer	Female	1 years	Chic	Hiphop
P7	Brand Designer	Male	4 years	Hiphop	Chic
P8	Freelancer Designer	Female	1 years	Feminine	Hiphop
P9	Freelancer Designer	Male	2 years	Hiphop	Sporty
P10	Brand Designer	Female	5 years	Feminine	Sporty

5.3 Study Procedure

The study lasted approximately 90 minutes per participant and followed a within-subjects design, with each participant completing two rounds of moodboard creation and sketch tasks in two different system environments: *QuerySwitch* and the baseline (ChatGPT with DALL-E 3). To simulate the early-stage vagueness that fashion designers typically experience, participants were instructed to begin with a broad style concept. To ensure participants engaged with the task from a state of vagueness, style keywords were randomly assigned from the following set: feminine, vintage, sporty, chic, and hip-hop. These keywords were selected through informal discussions with three newly recruited fashion designers, who confirmed that each keyword represented a distinct and non-overlapping design direction. Each participant received a different style keyword for the *QuerySwitch* and baseline sessions to prevent exposure to detailed designs in one session from influencing performance in the other (Table 2).

Each round began with a two-minute tutorial to familiarize participants with the assigned tool. Based on the assigned style keywords, participants were free to formulate their own initial prompt in each round. Participants then created their moodboards, expanded sketch ideas based on the moodboards, and produced final sketches on paper using a pen. After each round, they completed a post-task survey. Upon finishing both rounds, a 15-minute semi-structured interview was conducted to capture participants' reflections on the differences between the two systems and the impact of the tools on their creative process. To facilitate recall, participants were shown screenshots of their interactions with each system. All interviews were audio-recorded and transcribed for analysis.

To avoid expectation effects, we implemented two procedures: (1) the order of tool use was counterbalanced across participants and (2) a 10-minute break was provided between the rounds, following prior studies [5, 6]. Additional rest time was provided upon participants' requests.

5.4 Measures

To evaluate whether *QuerySwitch* effectively supports the balancing of vagueness and improves user experience with LLMs in design, participants completed post-task surveys for both *QuerySwitch* and the baseline system. We measured the system's usability with the USE questionnaire [43], covering items on effectiveness, productivity, usefulness, control over activities, ease of accomplishment, time savings, ability to meet needs, and expected performance. In addition, we included five questions to assess participants' self-perceived experience with the AI system [76]. The survey also included the NASA-TLX questionnaire [19]. All items used a 7-point Likert scale, ranging from (1) strongly disagree to (7) strongly agree [37] (A list of survey questions are listed in Appendix).A paired t-test was conducted, with the normality assumption met.

Given that imbalanced vagueness can lead to either continuous shifts away from the core concept or design fixation, we evaluated *consistency* and *complexity* using established computational measures. To assess consistency, we used BERT [42] and CLIP [52] embeddings, which are widely applied in design research to evaluate semantic relevance [27, 66]. Pairwise distances were computed, with lower values indicating higher semantic consistency. To assess complexity, we applied Shannon entropy [38] for text and edge map entropy [75] for images, both validated as indicators of diversity in design outputs [9, 18, 35]. Higher entropy values reflect broader and more evenly distributed ideas.

Specifically, for text queries, we measured pairwise distances between user input queries with BERT embeddings and computed Shannon entropy for each query. For visual outputs, we measured pairwise distances between images with CLIP embeddings and calculated edge map entropy to capture structural diversity. Independent-samples t-test was used when normality was met, otherwise Mann-Whitney U test was applied. Beyond grouped analyses of query and image outputs, we conducted a phase-level analysis to assess QuerySwitch's robustness across the four design stages.

For the qualitative component, we conducted in-depth semi-structured interviews to examine participants' perceptions of how *QuerySwitch* differed from the baseline in supporting the creative process. The interviews were structured around four themes: (1) How were the outputs of *QuerySwitch* and the baseline used in generating moodboards and sketches? (2) How did the input methods of each system influence the usability of generative tools in creative work? (3) Were the two systems applicable to iterative divergent-convergent process? and (4) How can each system be applied in real-world design processes? Three authors of this paper independently coded the interview transcripts and iteratively compared results, resolving disagreements through repeated discussions. Inter-coder reliability was high, with Cohen's Kappa values above 0.78 for all categories (max = 0.86, average = 0.81). We denote participant quotes using "PX," referring to "Participant Number X."

6 Results

6.1 (RQ1) Regulating Consistency and Diversity in Divergence and Convergence

6.1.1 Hierarchical Keywords: Maintaining Consistency while Exploring Diverse Ideas During Divergence. Quantitative and qualitative analyses indicate that hierarchical keywords helped participants balance vagueness by maintaining alignment with the main concept while supporting diverse exploration (Table 3). Because participants frequently consulted the hierarchical keywords before entering queries, *QuerySwitch* produced inputs with significantly higher complexity than the baseline (Complexity, t = -5.17, p < 0.001), indicating greater diversity. Phase-level analysis showed that both the Moodboard (mean = 4.58, std = 1.31) and Sketch parallel queries (mean = 4.11, std = 0.90) exhibited Complexity levels comparable to the overall *QuerySwitch* queries (mean = 3.49, std = 1.78), demonstrating robustness of *QuerySwitch* across phases. The Abstract Query yielded lower complexity, due to its inherently short form (mean = 1.11, std = 0.88). Queries through *QuerySwitch* also demonstrated higher semantic similarity (Consistency, U = 3007.50, p < 0.001), indicating more coherent concept alignment than the baseline. Analyzing by query type, both the Moodboard (mean = 13.18, std = 8.66) and the Sketch parallel queries (mean = 9.69, std = 7.30) showed Consistency levels comparable to the overall *QuerySwitch* queries (mean = 13.16, std = 3.08), indicating cross-phase stability. Because consistency is computed based on pairwise similarity, it was calculated only when at least two queries existed within a phase. Accordingly, abstract queries, typically used only once per phase, were excluded from the analysis.

Participants' reflections reinforced these quantitative results. Several noted that hierarchical keywords helped them preserve vague mental images coherent with their concept. *"At first, I only imagined 'feminine' as a dress [...] but since the system suggested detailed styles such as 'romantic feminine' and 'chic feminine,' along with specific items for each style, I began to think of various extensions—like transforming a skirt or matching it with new elements"* (P8, Figure 5). Another participant recognized how the system broadened a biased initial mental image, while still maintaining the main concept. *"When I thought of vintage, I only imagined bold patterns. But after the system suggested vintage elements such as 'straight-fit skirt,' 'clutch,' and 'apron,' I realized my view was too narrow. So I began to think of more inclusive vintage silhouettes"* (P5). These findings illustrate how hierarchical keywords supported participants in sustaining rough but consistent mental imagery—balancing vagueness by promoting both exploration and coherence.

Some participants indicated that hierarchical keywords could realign ideas with the main concept when divergent thinking became overly broad. For example, one participant noted that while focusing on novelty, her moodboard drifted from the main concept, which she recognized through the hierarchical keywords. *"I wanted to create a new chic-mannish concept, so I kept combining different elements in the moodboard [...] Later, when I entered the moodboard and reviewed the derived keywords, I realized it leaned too heavily toward suit style. So I adjusted my thinking by emphasizing chic elements and re-entered the query"* (P1). These findings suggest that when mental imagery drifts from the concept and vagueness grows excessive, hierarchical keywords can help prevent such divergence and effectively support the balance of vagueness.

Hierarchical keywords were also seen as useful in collaborative contexts, where differing initial mental images often cause collective deviation from the concept. *"Because everyone thinks differently, it always ends up all over the place. So we usually write the concept on the first line of a Word document, then paste in the ideas we each found and spend a long time reviewing them together, gradually removing the ones that don't fit the concept. [...] At that point, these keywords could help us quickly filter out ideas that deviate too much"* (P2). In this way, hierarchical keywords can support not only individual workflows but also group collaboration by reducing inconsistencies and maintaining alignment with the central concept.

6.1.2 Combinational Images: Supporting Flexibility and Preventing Fixation during Convergence. We found that combinational images supported participants in balancing vagueness by enabling them to concretize their concepts while still considering diverse outcomes (Table 3). Compared to the baseline, *QuerySwitch* generated images with higher semantic similarity (Consistency, U = 7.50, p = 0.001). Phase-level analysis showed that both the Moodboard (mean = 0.52, std = 0.16) and the Sketch combination images (mean = 0.72, std = 0.09) maintained Consistency levels similar to the overall *QuerySwitch* image output (mean = 0.71, std = 0.43), indicating robustness of *QuerySwitch* across phases. Consistency was computed only when at least two images were generated within each phase, due to its reliance on pairwise distance. Combination images also produced more complex outputs compared to the baseline (Complexity, U = 3007.50, p < 0.001), demonstrating greater diversity. When examined by phase, both the Moodboard (mean = 0.58, std = 0.03) and Sketch combination images (mean = 0.55, std = 0.05) maintained Complexity levels comparable to the overall *QuerySwitch* image output (mean = 0.56, std = 0.04), indicating cross-phase stability.

Participants also reported that these combinations helped them avoid fixation and foster more fluid thinking. *"After seeing the keywords, I decided to create a feminine moodboard with floral patterns, ruffles, and dresses, and when I entered the query, it also generated fresh items like a lace corsage and a floral cardigan. [...] I was able to complete a moodboard that was more distinctive than the feminine style I had initially imagined"* (P8, Figure 5). Another participant noted that real-world design practices are highly constrained, which often leads to idea fixation, and suggested that combinational images could support more flexible thinking. *"To maintain the target customer group, we can never deviate from the brand concept, and sometimes we even have to refer to designs that sold well just before. So when sketching, I often get stuck with a single idea. In those moments, I believe this system could provide insights such as, 'Oh, I can slightly modify this element,' which would be very helpful"* (P3). These reflections suggest that combinational images effectively help prevent

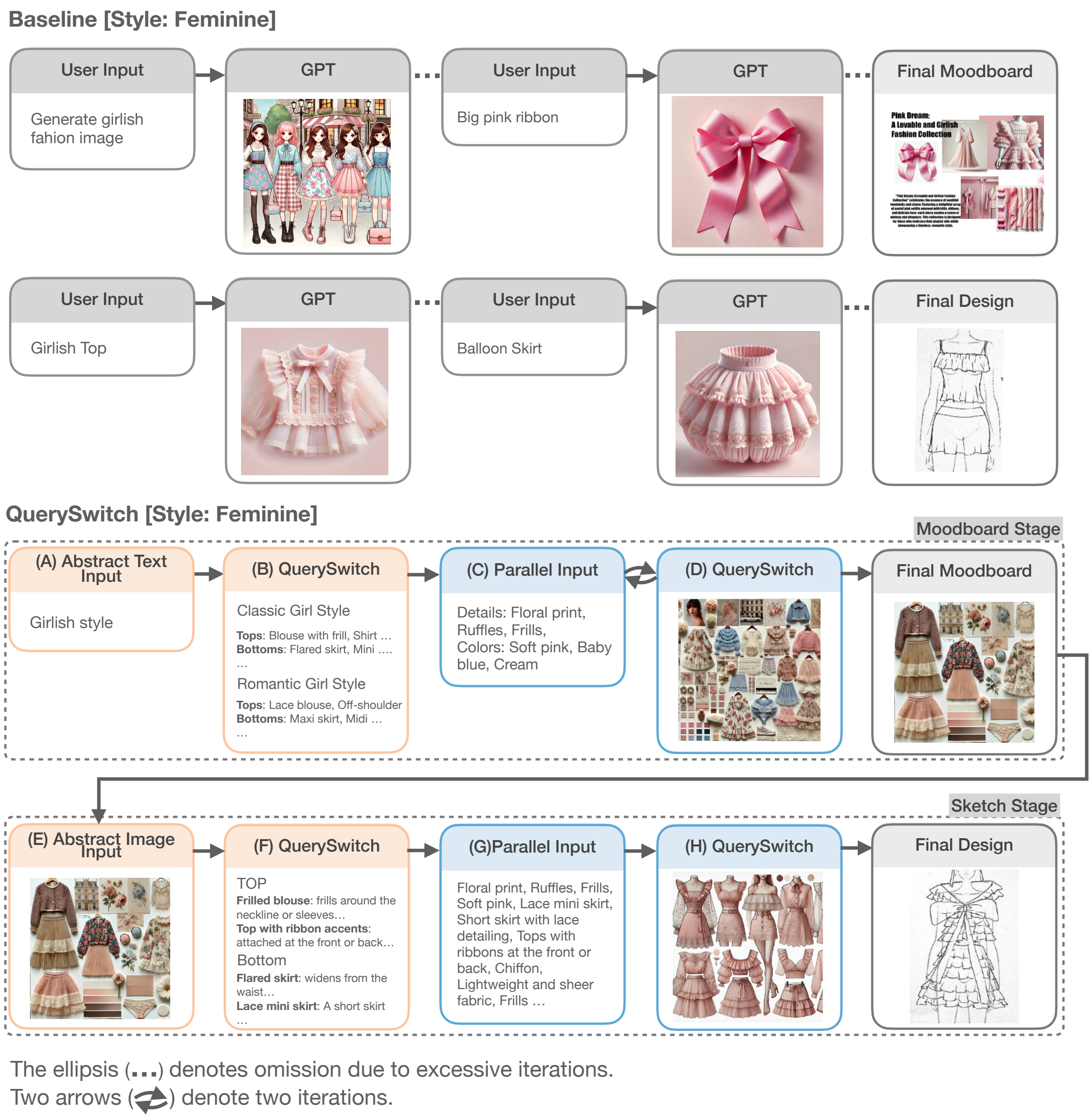

Figure 5: Comparison of design outcomes generated with *QuerySwitch* and the baseline system. *QuerySwitch* supports the management of vagueness by enabling consistent design direction while allowing flexible ideation across stages. In contrast, the baseline system often leads to redundant iterations and exhibits greater uncontrolled variation.

Table 3: Results of independent-samples t-tests and Mann–Whitney U tests comparing text query consistency and complexity, as well as image output consistency and complexity, between *QuerySwitch* and the baseline system ($^{**}p < 0.01$, $^{***}p < 0.001$). d indicates Cohen's d (effect size) for t-tests, and r indicates the rank biserial correlation (effect size) for Mann–Whitney U tests.

		QuerySwitch		Baseline		Statistics	
		mean	std	mean	std	$t(18)/U$	d/r
Text Query	Consistency (BERT Pairwise Distances)	13.16	3.08	20.12	2.92	$t(18)$ = -5.17***	-2.31
	Complexity (Shannon Entropy)	3.49	1.78	2.41	0.90	U = 3007.50***	0.42
Image Output	Consistency (CLIP Pairwise Distances)	0.71	0.43	1.54	0.28	U = 7.50**	0.85
	Complexity (Edge Map Entropy)	0.56	0.04	0.48	0.10	U = 1421.00***	0.53

Table 4: Results of paired t-tests comparing user experience and self-perceived experience with the AI between *QuerySwitch* and the baseline system ($^{\dagger}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$). d indicates Cohen's d (effect size) for t-tests.

		QuerySwitch		Baseline		Statistics	
		mean	std	mean	std	$t(18)$	d
USE Questionnaire	Effective	6.5	0.97	4.6	1.65	$t(18) = 2.75^{*}$	$d = 0.87$
	Productive	6.3	1.25	4.8	1.75	$t(18) = 2.29^{*}$	$d = 0.73$
	Useful	6.4	0.84	5.0	1.33	$t(18) = 2.59^{*}$	$d = 0.82$
	Control Activities	6.3	0.82	3.9	1.66	$t(18) = 4.43^{**}$	$d = 1.40$
	Accomplish Easier	6.5	0.71	4.4	1.96	$t(18) = 3.19^{*}$	$d = 1.01$
	Save Time	6.6	1.27	4.9	2.03	$t(18) = 2.23^{\dagger}$	$d = 0.70$
	Meet Needs	6.2	0.92	4.2	1.62	$t(18) = 3.16^{*}$	$d = 1.00$
	Do Expected	4.9	1.66	3.2	1.48	$t(18) = 2.68^{*}$	$d = 0.85$
Self-perceived Experience	Match Goal	6.2	1.14	4.2	1.69	$t(18) = 2.54^{*}$	$d = 0.80$
	Think Through	6.6	0.67	4.4	2.12	$t(18) = 3.16^{*}$	$d = 1.00$
	Transparent	6.2	0.92	5.3	1.34	$t(18) = 2.38^{*}$	$d = 0.75$
	Controllable	6.3	1.16	4.9	1.73	$t(18) = 2.41^{*}$	$d = 0.76$
	Collaborative	6.5	1.08	4.2	1.62	$t(18) = 3.54^{**}$	$d = 1.12$

premature convergence during the sketch stage, helping designers preserve both direction and openness.

Some participants anticipated benefits in collaborative refinement. In team settings, moodboards often lead to homogenized sketches, as designers converge on similar outputs for consistency. Combinational images were expected to counteract this by offering variations within overlapping elements. *"We work in the same brand, so inevitably we're exposed to the same images, and our designs often turn out similar. But if we input the overlapping parts here, it gives slightly different suggestions for the same elements, so we can look at them together and each modify them in our own way. That would be really helpful"* (P10). By presenting diverse interpretations of shared inputs, combinational images can enhance the fluidity of outcomes and support balanced vagueness even in collaborative design processes.

At the same time, limitations emerged. One participant noted the lack of information on current fashion trends. *"The biggest strength is that it can generate images that don't exist in the real world, but I don't think it reflects trends. I still need to supplement information through sources like Vogue.com [where runway images are uploaded], or Instagram"* (P6). Another participant mentioned that cultural contexts were not represented. *"In our country, women don't usually wear clothes with a lot of exposure. But in the generated images, sometimes the neckline was far too deep. I still need to rely on experiential knowledge I gain by visiting stores or observing people on the street"* (P7). These concerns suggest that while *QuerySwitch* supports conceptual exploration, it should be complemented by trend database and designers' experiential knowledge to ensure socio-cultural relevance.

6.2 (RQ2) Efficiency and User Agency within the Design Process

6.2.1 Minimizing Query Adjustment Effort through Abstract and Parallel Input. Participants reported that *QuerySwitch* allowed them to focus on design work without investing much attention in query adjustment. They explained that the system produced sufficiently relevant information even they relied only on abstract or parallel queries. *"When my thoughts were abstract, I could simply enter the concept and the system suggested a wide range of ideas. When my thoughts became somewhat more concrete, I just pasted fashion elements into the query, and it provided information that helped me refine my ideas. Without the need to fine-tune the query, I could spend more time sketching"* (P10). In contrast, the baseline system struggled to generate appropriate results from vague inputs, forcing participants to repeatedly refine their queries. *"When I asked it to generate a vivid hip-hop style, it gave me folk-style images. After considerable effort adding hip-hop fashion elements, the results suddenly converged into overly uniform hip-hop images lacking diversity"* (P6).

Survey results confirmed these differences. On the USE questionnaire, participants rated *QuerySwitch* more effective ($t = 2.75$, $p = 0.022$), productive ($t = 2.29$, $p = 0.048$), useful ($t = 2.59$, $p = 0.029$), adjustable to their desired way of working ($t = 4.43$, $p = 0.002$), faster in achieving results ($t = 3.19$, $p = 0.011$), responsive to their needs ($t = 3.16$, $p = 0.012$), and successful in delivering expected outcomes ($t = 2.68$, $p = 0.025$) than the baseline, with marginal differences in time-saving ($t = 2.23$, $p = 0.052$) (Table 4). These findings demonstrate that *QuerySwitch* reduces the need for iterative query adjustments by providing suitable information from vague expressions, allowing designers to maintain focus on their work.

6.2.2 Workflow Customization and User Agency through Shifting Input-Output Modes. From participants' usage patterns, we found that *QuerySwitch* could be flexibly adapted to each designer's preferred working style. Participants who primarily worked on apparel tended to proceed through the system's four query flows (A–C–E–G) sequentially. However, at the sketching interface, two distinct query types emerged: participants who worked across multiple categories mainly used the (G-1) type queries, whereas those who focused on a single fixed category primarily used the (G-2)

Table 5: Results of paired t-tests comparing NASA-TLX total score and subcomponents between *QuerySwitch* and the baseline system ($^{\dagger}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$). d indicates Cohen's d (effect size) for t-tests.

		QuerySwitch		Baseline		Statistics	
		mean	std	mean	std	$t(18)$	d
	Overall Score	18.1	14.94	35.4	18.21	$t(18) = -2.55^{*}$	$d = -0.81$
NASA-TLX	Mental	21.0	26.23	37.5	29.37	$t(18) = -1.69$	$d = -0.53$
	Physical	11.0	24.47	15.5	21.27	$t(18) = -0.46$	$d = -0.15$
	Temporal	17.0	24.06	37.0	28.98	$t(18) = -2.41^{*}$	$d = -0.76$
	Effort	24.0	18.07	37.0	13.78	$t(18) = -1.90^{\dagger}$	$d = -0.60$
	Performance	83.0	15.13	51.0	23.78	$t(18) = 3.42^{**}$	$d = 1.08$
	Frustration	18.5	19.44	36.5	32.75	$t(18) = -1.44$	$d = -0.45$

type queries. In addition, some participants repeatedly leveraged the system depending on their brand's core product range. *"Our brand focuses not only on women's clothing but also on sneakers. After completing a dress sketch, I could re-upload the reference images and receive keyword suggestions to continue into sneaker design"* (P10).

The Self-perceived Experience evaluation further supported these qualitative findings. *QuerySwitch* received significantly higher ratings than the baseline on match goal (t = 2.54, p = 0.032), think through (t = 3.16, p = 0.012), transparent (t = 2.38, p = 0.041), controllable (t = 2.41, p = 0.039), and collaborative (t = 3.54, p = 0.006) (Table 4). These results demonstrate that participants were able to actively adapt *QuerySwitch* to their design processes and exercise agency in its use.

Participants attributed this flexibility to the ability to choose interfaces aligned with their design phase. *"If I wanted to specify a styling direction, I could simply click the moodboard interface, and if I wanted to work on more detailed design tasks, I could click the sketch interface. Because I only had to choose the appropriate screen depending on the situation, it was much more convenient to adjust"* (P9). In contrast, the baseline forced participants to adapt to the system rather than maintain their own workflow. *"This system [QuerySwitch] had phases, but the baseline did not. So I kept making arbitrary adjustments, which led to disorganized results, and in the end, it only showed representative images. Instead of fitting the system to my process, I had to give up and adapt my design to the results"* (P4, Figure 5). This contrast demonstrates that *QuerySwitch* 's ability to separate input-output modes by design phase enabled participants to align the system with their creative workflow and preserve agency in the design process.

6.2.3 Higher Design Achievement with Equivalent Effort. NASA-TLX results confirmed that *QuerySwitch* enabled higher design outcomes without increasing user workload (Table 5). The overall NASA-TLX score was significantly lower for *QuerySwitch* (t = -2.55, p = 0.031). Among the six subscales, Temporal (t = -2.41, p = 0.039) and Performance (t = 3.42, p = 0.008) showed a notably positive evaluation, and Effort showed a marginally positive difference (t = -1.90, p = 0.090). In contrast, for Mental (t = -1.69, p = 0.126), Physical (t = -0.46, p = 0.655) and Frustration (t = -1.44, p = 0.185), the differences were not statistically significant, though *QuerySwitch* still received slightly more favorable mean ratings.

Participants explained that while the baseline allowed free-form input, it often required repeated adjustments due to inappropriate results. In contrast, *QuerySwitch* involved slightly more procedural steps but consistently delivered useful information that led to higher satisfaction with the outcomes. *"DALL-E allows freer input than this [QuerySwitch], which is an advantage. But often the results weren't appropriate, so I had to make several adjustments. [...] On the other hand, this [QuerySwitch] had a few steps compared to DALL-E, but it gave me much more useful information, so I could create designs I was more satisfied with"* (P1). These findings indicate that *QuerySwitch* yielded greater design accomplishment by providing more balanced and relevant information.

7 Discussion

7.1 The Importance of Identifying Structures of Idea Expansion from the Core Concept

During divergence participants managed vagueness effectively by using hierarchical keywords to stay anchored to the main concept (Section 6.1.1). Moreover, when ideas became overly fluid and vagueness expanded, the hierarchical structure realigned participants' thinking with the core concept, helping to balance vagueness. This anchoring effect can be attributed to the structure's ability to provide flexibility within the boundaries of a core frame. For example, five sub-styles derived from core style-related keywords expanded into nine categories each, and moodboard images followed a similar expansion. This structure allowed participants to loosely maintain concept-related mental images, enabling exploration without excessive deviation from the concept.

Therefore, to balance vagueness during divergence, it is important to establish a structure for expanding ideas from a core concept and support exploration through branching extensions. This approach resembles methods where LLMs immediately generate new keywords with each user input [5, 6, 50, 64, 66]. However, it differs in two key aspects: expansion remains anchored exclusively to the core concept, and the expansion structure is tailored to that concept. This enables them to sustain a rough mental image, preserving conceptual coherence while still exploring broadly.

There are two possible implementation strategies. First, developers can embed such predefined structures into the system. For example, in domains where the core concept is style (e.g., fashion,

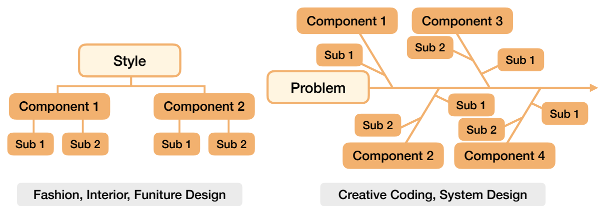

Figure 6: An example of idea expansion structures based on a core concept. Hierarchical structures are well suited to domains where style is central, such as fashion, interior, and furniture design, whereas fishbone structures are more appropriate for domains focused on problem-solving, such as creative coding and system design.

interior, or furniture design), a hierarchical structure can be applied, whereas in problem-solving domains (e.g., creative coding or system design), a fishbone structure may better fit the workflow (Figure 6). Second, the system allows users to select and adopt a structure. When a concept is entered, the interface could suggest options such as hierarchical, fishbone, or radial expansions. In both cases, flexibility is preserved within the boundaries of a core concept, enabling users to maintain conceptually consistent yet vague mental images and to prevent excessive divergence.

7.2 The Importance of Combining Basic Elements with Concrete Ideas

During convergence, participants balanced vagueness by envisioning concrete ideas while keeping their application open and flexible (Section 6.1.2). This was achieved through a combinational structure, which prevents premature convergence on a single idea and enables participants to consider multiple ways of implementing ideas. For example, participants' concrete ideas in their queries (e.g., pink, lace, oversized) were combined with basic style elements (e.g., color palette, silhouette, fabric) in the moodboard interface, or with clothing components (e.g., tops, bottoms, dresses) in the sketch interface. Rather than presenting a single optimal solution, the system diversified how these inputs could be realized, enabling participants to sustain vague mental images until the outcome was finalized and explore a wider range of creative results.

Therefore, to balance vagueness during convergence, it is important to identify the basic components that constitute an outcome and enable diverse ways of combining them with user input. Prior studies have highlighted the importance of producing multiple options at this stage [6, 25], but our findings emphasize not just numerical diversity, but structural diversity—the different ways user ideas can be integrated into outcomes. This ensures that users can envision concrete ideas while retaining the flexibility necessary to avoid design fixation.

One design implication is to predefine domain-specific basic elements and embed them into the system so that combinations can be automatically generated. For example, in the illustration domain, elements such as composition, color, line and shape, and texture could be predefined. When users specify ideas (e.g., pastel tones, dynamic poses, emphasis lines), the system could combine these with the basic elements to produce a variety of visual outputs. Alternatively, interfaces could allow users to link basic elements and detailed ideas as nodes, with the system generating images from these combinations. Both approaches preserve flexibility in how ideas are concretized, helping users balance the vagueness during convergence and avoid premature fixation.

7.3 The Need to Preserve Users' Expressive Styles in Queries

Participants were able to focus on their design work without expending significant effort on query adjustments (Section 6.2.1). This was because the system not only generated appropriate information but also allowed participants to retain the expressive styles that naturally emerge from vagueness in their design processes, eliminating the need for additional query refinement. Specifically, abstract queries—such as keywords for fashion style—and parallel queries—where multiple detailed ideas were listed—were sufficient to generate hierarchical keywords and combinational images suitable for the divergent-convergent process. By contrast, requiring designers to precisely specify the intended outcomes would undermine the vagueness-driven nature of design practice. By supporting abstract and parallel queries, and generating structural outputs accordingly, the system improved usability without disrupting core creative practices.

To maximize the usability of LLMs in vagueness-driven design, it is therefore essential to preserve vague expressions in user queries during both divergent and convergent phases. Forms of vagueness will differ across domains. For example, in interior design, abstract elements such as style may appear in divergent-stage queries, while more concrete interior ideas emerge during convergence. In novel writing, genres or temporal settings may characterize divergent queries, whereas plot developments or character traits may appear in convergent queries. Once such input styles are identified, the generation methods outlined in Sections 7.1 and 7.2 can be applied to provide stage-appropriate outputs under vague input conditions.

Supporting vague queries alleviates the burden of query refinement, which impose a high cognitive load [63]. This is especially important in vagueness-driven processes, where design intentions often cannot be fully articulated during the creative process. Preserving vagueness while still producing structured outputs ensures both usability and creative alignment in LLM-based systems.

7.4 The Need for Automated Context Recognition for Smooth Divergence–Convergence Shifts

Since *QuerySwitch* distinguished between query–output modes for each design phase, participants only needed to switch queries to obtain the required information, allowing them to adapt the system according to their own divergent–convergent workflow (Section 6.2.2). Prior studies have also demonstrated that distinguishing queries by output type can facilitate smooth transitions to task-relevant results [51, 78]. We suggest that communication with LLMs could be further enhanced if open-ended queries were automatically interpreted by the system—recognizing context and generating hierarchical and combinational information without requiring explicit mode-switching.

Therefore, to enhance usability in iterative divergent–convergent process, LLM systems need mechanisms that can automatically recognize context, even from a single open-ended query. Prior research has emphasized the importance of situating LLMs within users' divergent–convergent contexts [79]. We emphasize that this capability is necessary even at the level of query interpretation. Recent studies explores ways for LLMs to automatically infer user intentions from natural conversational input with implicit signals [58, 82].

Extending such approaches to creative domains is a critical next step. One method is to establish a threshold between vagueness and concreteness in queries to classify inputs as divergence or convergence. For example, as described in Section 7.3, queries dominated by abstract concepts (e.g., style) can be mapped to the moodboard divergence phase, whereas queries focused on detailed fashion elements can be mapped to the sketch convergence phase. If a query lacks sufficient information to determine its stage, the system could prompt the user with clarifying questions (e.g., "It seems you would like to expand your ideas, what is your main concept?") In addition, the system could also guide stage transitions. For example, if hierarchical keywords have been generated repeatedly, the system might infer that the user has sufficiently explored divergence and suggest "You may now select some keywords." Embedding such adaptive guidance into natural, open-ended interactions would heighten users' awareness of their current stage and clarify how the system supports their workflow. Ultimately, automated context recognition can help users more effectively leverage LLMs across iterative divergence–convergence cycles, reducing friction in query formulation while preserving the exploratory and flexible nature of creative workflows.

7.5 Generalizability to Domains with Fluid Design Elements

Although participants noted that hierarchical keywords and combinational images supported the management of vagueness (Sections 6.1.1 and 6.1.2), in domains where the elements constituting the creative work are fluid or not yet well-defined, generating structured ideas by pre-fixing the elements may constrain creative thinking. This is because designers in these domains often explore a broad and unpredictable range of possible elements. For example, while fashion designers typically work with within a relatively fixed set of elements (e.g., tops, bottoms, textures), contemporary artists may draw on a far more diverse and unexpected range of elements (e.g., natural objects, sound, lighting) [59]. In such cases, the breadth of exploration may be hindered when idea generation is constrained by a pre-defined and rigid set of elements.

Nevertheless, maintaining an appropriate level of vagueness is a key strategy across diverse creative domains [15, 44]. Thus, structured exploration can still be a viable approach even when the constituent elements are fluid. To be effective, however, it is important for users to directly specify the elements that will constitute the basis of the structured ideas. Users may explicitly provide the elements they wish to explore, or when the elements have not yet been determined, the system can first suggest a range of plausible candidates based on a conceptual input. For example, in domains such as installation or performance art, entering a high-level concept such as "dreamlike atmosphere" could prompt the system to suggest potential elements (e.g., lighting types, sound textures, movement trajectories). Users can then select or refine these suggestions, after which the system generates structured ideas grounded in the user-specified elements. This iterative, user-driven specification process enables the system to maintain an appropriate balance of vagueness even when design elements are fluid and evolving.

It is also important to position such tools as one option among a broader ecosystem of creative resources, rather than as the sole alternative. For example, contemporary artists often draw inspiration from unpredictable external sources such as magazines or news articles; once a particular idea warrants deeper exploration, they may incorporate structured exploration tools to further develop that element. Even in domains with relatively well-defined constituent elements (e.g., furniture design, interior design), there are stages in which the elements become more fluid. In such moments, other tools (e.g., material reference websites, 3D rendering software) can serve as supplementary resources. For example, furniture designers may rely primarily on structured exploration during ideation (using tools such as *QuerySwitch*), but when transitioning to real-world considerations (e.g., materials, physical properties, and manufacturing constraints), they may integrate additional tools to broaden the range of elements. This process can iterate until the outcome becomes satisfactory. These observations highlight the need for a flexible tool-integration strategy that adapts to the level of element fluidity. Future work should explore not only user-centered approaches for cultivating such awareness but also system-level mechanisms that allow multiple creative tools to operate in a complementary and adaptive manner.

7.6 Limitation and Future Work

While our study offers many valuable insights, there are several limitations that need to be addressed in future research.

First, although combinational images effectively supported vagueness management, some participants noted that such images did not sufficiently reflect social trends. Prior research similarly showed that LLMs often struggle to incorporate updated information or capture evolving social dynamics [48]. This suggests that designers should not rely solely on LLM-generated information but instead complement it with socio-cultural data and trend resources. Future research could explore integrating LLMs with dynamic datasets such as street fashion images and Vogue Runway collections to better situate outputs in real-world contexts.

Second, the perspectives of ten participants may not fully capture the diversity of practices across the fashion industry, which spans from luxury to SPA brands. Expanding participant pool in scale and variety will allow for a more comprehensive understanding of how vagueness is managed across different subdomains and contexts. Nonetheless, given that the importance of balancing vagueness is widely recognized in design, our study provides meaningful implications for aligning LLM-based tools with creative workflows.

Third, the limited duration of our study may not reflect the full scope of the fashion design process, which often spans over six months to a year. Longitudinal investigations are needed to examine how vagueness is sustained, negotiated, and transformed across extended design cycles. Such research would provide deeper insights into how LLM-based systems can support designers across iterative and evolving workflows.

Finally, our system relies primarily on prompting as the core interaction mechanism. While this approach may not encompass all possible forms of creative engagement, prompting has been increasingly recognized as a flexible and effective means of facilitating interaction with generative AI [64–66]. Our contribution lies in structuring and tailoring prompting strategies to align with domain-specific creative practices. Building on this foundation, future research should investigate complementary modalities, such as sketch-based, multimodal, or adaptive interaction, to broaden creative possibilities and better accommodate diverse workflows.

8 Conclusion

This study examined the role of vagueness in creative design, highlighting its importance in fostering diverse interpretations and preventing early fixation. Effectively managing vagueness, however, requires structured support throughout the iterative divergent-convergent process—an area where existing LLM-based tools have proven limited. To address this gap, we developed *QuerySwitch*, an interactive prototype that allowed users to switch between two query-output modes to better balance vagueness. Our user study with professional fashion designers showed that *QuerySwitch* effectively supported vagueness management and improved the usability of LLMs in creative workflows. By accommodating both abstract and parallel inputs, the system aligned with different stages of the design process and enabled more personalized, context-aware ideation. These findings underscore the value of combining structured exploration mechanisms with support for users' expressive query styles, with promising implications for future LLM-based creativity support tools across a range of domains.

Acknowledgments

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.RS-2020-II201373, Artificial Intelligence Graduate School Program (Hanyang University)) and the Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea government(MOTIE) (No. RS-2025-02633871, Development of automated data pre-processing technology for industrial AI)

References

[1] Jimmy Abualdenien and André Borrmann. 2020. Vagueness visualization in building models across different design stages. *Advanced Engineering Informatics* 45 (2020), 101107.
[2] James Betker, Gabriel Goh, Li Jing, Tim Brooks, Jianfeng Wang, Linjie Li, Long Ouyang, Juntang Zhuang, Joyce Lee, Yufei Guo, et al. 2023. Improving image generation with better captions. *Computer Science. https://cdn. openai. com/papers/dall-e-3. pdf* 2, 3 (2023), 8.
[3] Stephen Brade, Bryan Wang, Mauricio Sousa, Sageev Oore, and Tovi Grossman. 2023. Promptify: Text-to-image generation through interactive prompt exploration with large language models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–14.
[4] John M Carrol. 1999. Five reasons for scenario-based design. In *Proceedings of the 32nd annual hawaii international conference on systems sciences. 1999. hicss-32. abstracts and cd-rom of full papers*. IEEE, 11–pp.
[5] Liuqing Chen, Qianzhi Jing, Yixin Tsang, Qianyi Wang, Ruocong Liu, Duowei Xia, Yunzhan Zhou, and Lingyun Sun. 2024. AutoSpark: Supporting Automobile Appearance Design Ideation with Kansei Engineering and Generative AI. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–19.
[6] 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 2024 CHI Conference on Human Factors in Computing Systems*. 1–25.
[7] Prafulla Dhariwal and Alexander Nichol. 2021. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems* 34 (2021), 8780–8794.
[8] Yingjing Duan and Jie Zhang. 2022. A Novel AI-Based Visual Stimuli Generation Approach for Environment Concept Design. *Computational Intelligence and Neuroscience* 2022, 1 (2022), 8015492.
[9] Ondřej Dušek, Jekaterina Novikova, and Verena Rieser. 2020. Evaluating the state-of-the-art of end-to-end natural language generation: The e2e nlg challenge. *Computer Speech & Language* 59 (2020), 123–156.
[10] Ronald A Finke, Thomas B Ward, and Steven M Smith. 1996. *Creative cognition: Theory, research, and applications*. MIT press.
[11] Angie Finn and Lee Finn. 2010. Uncertainty and innovation in fashion design. In *Global Fashion: Innovative and Creative Contexts: Proceedings of the 1st International Conference on Fashion*. Centre of Research on Economics, Population and Society-CEPESE, 1–26.
[12] Jonathan Fish and Stephen Scrivener. 1990. Amplifying the mind's eye: sketching and visual cognition. *Leonardo* 23, 1 (1990), 117–126.
[13] Liane Gabora. 2005. Creative thought as a non Darwinian evolutionary process. *The Journal of Creative Behavior* 39, 4 (2005), 262–283.
[14] Liane Gabora and Adam Saab. 2013. Creative interference and states of potentiality in analogy problem solving. *arXiv preprint arXiv:1308.4241* (2013).
[15] William W Gaver, Jacob Beaver, and Steve Benford. 2003. Ambiguity as a resource for design. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 233–240.
[16] Vinod Goel. 2014. Creative brains: designing in the real world. *Frontiers in human neuroscience* 8 (2014), 241.
[17] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative adversarial networks. *Commun. ACM* 63, 11 (2020), 139–144.
[18] Maria Grebenkina, Anselm Brachmann, Marco Bertamini, Ali Kaduhm, and Christoph Redies. 2018. Edge-orientation entropy predicts preference for diverse types of man-made images. *Frontiers in Neuroscience* 12 (2018), 678.
[19] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.
[20] Sven Hawer, Alexander Schönmann, and Gunther Reinhart. 2018. Guideline for the Classification and Modelling of Uncertainty and Fuzziness. *Procedia Cirp* 67 (2018), 52–57.
[21] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems* 33 (2020), 6840–6851.
[22] Rong Huang, Haichuan Lin, Chuanzhang Chen, Kang Zhang, and Wei Zeng. 2024. Plantography: Incorporating iterative design process into generative artificial intelligence for landscape rendering. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–19.
[23] Saskia Jaarsveld and Thomas Lachmann. 2017. Intelligence and creativity in problem solving: the importance of test features in cognition research. *Frontiers in psychology* 8 (2017), 134.
[24] Sun Hee Jang, Byungkeun Oh, Sukil Hong, and Jinwoo Kim. 2019. The effect of ambiguous visual stimuli on creativity in design idea generation. *International Journal of Design Creativity and Innovation* 7, 1-2 (2019), 70–98.
[25] Youngseung Jeon, Seungwan Jin, Patrick C Shih, and Kyungsik Han. 2021. FashionQ: an ai-driven creativity support tool for facilitating ideation in fashion design. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–18.
[26] Christopher Kennedy. 2011. Ambiguity and vagueness: An overview. *Semantics: An international handbook of natural language meaning* 1 (2011), 507–535.
[27] 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.
[28] Myungjin Kim, Misun Joo, and Kyungsik Han. 2024. CoCoStyle: Mixed initiative co-creative system to support creative process of fashion design. *SoftwareX* 27 (2024), 101857.
[29] Myungjin Kim, Misun Joo, and Kyungsik Han. 2025. Understanding the differences in an AI-based creativity support tool between creativity types in fashion design. *International Journal of Human–Computer Interaction* (2025), 1–14.
[30] Diederik P Kingma, Max Welling, et al. 2013. Auto-encoding variational bayes.
[31] George J Klir. 1987. Where do we stand on measures of uncertainty, ambiguity, fuzziness, and the like? *Fuzzy sets and systems* 24, 2 (1987), 141–160.
[32] Hyung-Kwon Ko, Gwanmo Park, Hyeon Jeon, Jaemin Jo, Juho Kim, and Jinwook Seo. 2023. Large-scale text-to-image generation models for visual artists' creative works. In *Proceedings of the 28th international conference on intelligent user*

interfaces. 919–933.

[33] Jane M Lamb and M Jo Kallal. 1992. A conceptual framework for apparel design. *Clothing and Textiles Research Journal* 10, 2 (1992), 42–47.

[34] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.

[35] Seung Won Lee, Tae Hee Jo, Semin Jin, Jiin Choi, Kyungwon Yun, Sergio Bromberg, Seonghoon Ban, and Kyung Hoon Hyun. 2024. The impact of sketch-guided vs. prompt-guided 3D generative AIs on the design exploration process. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–18.

[36] Yu-Hsu Lee and Chun-Yao Chiu. 2023. The impact of AI text-to-image generator on product styling design. In *International Conference on Human-Computer Interaction*. Springer, 502–515.

[37] Rensis Likert. 1932. A technique for the measurement of attitudes. *Archives of psychology* (1932).

[38] Jianhua Lin. 1991. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information theory* 37, 1 (1991), 145–151.

[39] Vivian Liu and Lydia B Chilton. 2022. Design guidelines for prompt engineering text-to-image generative models. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–23.

[40] Vivian Liu, Han Qiao, and Lydia Chilton. 2022. Opal: Multimodal image generation for news illustration. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–17.

[41] Vivian Liu, Jo Vermeulen, George Fitzmaurice, and Justin Matejka. 2023. 3DALL-E: Integrating text-to-image AI in 3D design workflows. In *Proceedings of the 2023 ACM designing interactive systems conference*. 1955–1977.

[42] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692* (2019).

[43] Arnold M Lund. 2001. Measuring usability with the use questionnaire12. *Usability interface* 8, 2 (2001), 3–6.

[44] Mary-Anne Mace and Tony Ward. 2002. Modeling the creative process: A grounded theory analysis of creativity in the domain of art making. *Creativity research journal* 14, 2 (2002), 179–192.

[45] Kathryn McKelvey and Janine Munslow. 2011. *Fashion design: process, innovation and practice*. John Wiley & Sons.

[46] Edward Morse, Jean-Yves Dantan, Nabil Anwer, Rikard Söderberg, Giovanni Moroni, Ahmed Qureshi, Xiangqian Jiang, and Luc Mathieu. 2018. Tolerancing: Managing uncertainty from conceptual design to final product. *CIRP Annals* 67, 2 (2018), 695–717.

[47] Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, and Mark Chen. 2021. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. *arXiv preprint arXiv:2112.10741* (2021).

[48] Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, and Xindong Wu. 2024. Unifying large language models and knowledge graphs: A roadmap. *IEEE Transactions on Knowledge and Data Engineering* 36, 7 (2024), 3580–3599.

[49] Or Patashnik, Zongze Wu, Eli Shechtman, Daniel Cohen-Or, and Dani Lischinski. 2021. Styleclip: Text-driven manipulation of stylegan imagery. In *Proceedings of the IEEE/CVF international conference on computer vision*. 2085–2094.

[50] 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.

[51] Hua Xuan Qin, Guangzhi Zhu, Mingming Fan, and Pan Hui. 2025. Toward Personalizable AI Node Graph Creative Writing Support: Insights on Preferences for Generative AI Features and Information Presentation Across Story Writing Processes. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–30.

[52] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PmLR, 8748–8763.

[53] Mohammad Rajabalinejad and Christos Spitas. 2012. Incorporating uncertainty into the design management process. *Design Management Journal* 6, 1 (2012), 52–67.

[54] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. Zero-shot text-to-image generation. In *International conference on machine learning*. Pmlr, 8821–8831.

[55] Royi Rassin, Shauli Ravfogel, and Yoav Goldberg. 2022. DALLE-2 is seeing double: Flaws in word-to-concept mapping in Text2Image models. *arXiv preprint arXiv:2210.10606* (2022).

[56] Mark A Runco. 2007. Creativity: Theories and themes: Research, development, and practice. *(No Title)* (2007).

[57] 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.

[58] Jeonghoon Shim, Gyuhyeon Seo, Cheongsu Lim, and Yohan Jo. 2025. Tooldial: Multi-turn dialogue generation method for tool-augmented language models. *arXiv preprint arXiv:2503.00564* (2025).

[59] Terry E Smith. 2009. *What is contemporary art?* University of Chicago Press.

[60] Kihoon Son, DaEun Choi, Tae Soo Kim, Young-Ho Kim, and Juho Kim. 2024. Genquery: Supporting expressive visual search with generative models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–19.

[61] Amanda Spink, Dietmar Wolfram, Major BJ Jansen, and Tefko Saracevic. 2001. Searching the web: The public and their queries. *Journal of the American society for information science and technology* 52, 3 (2001), 226–234.

[62] Martin Stacey and Claudia Eckert. 2003. Against ambiguity. *Computer Supported Cooperative Work (CSCW)* 12 (2003), 153–183.

[63] Hari Subramonyam, Roy Pea, Christopher Pondoc, Maneesh Agrawala, and Colleen Seifert. 2024. Bridging the gulf of envisioning: Cognitive challenges in prompt based interactions with llms. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–19.

[64] 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*. 1–26.

[65] Zhida Sun, Zhenyao Zhang, Yue Zhang, Min Lu, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. 2025. Creative Blends of Visual Concepts. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–17.

[66] Sirui Tao, Ivan Liang, Cindy Peng, Zhiqing Wang, Srishti Palani, and Steven P Dow. 2025. DesignWeaver: Dimensional Scaffolding for Text-to-Image Product Design. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–26.

[67] Winger SW Tseng and Linden J Ball. 2010. How uncertainty helps sketch interpretation in a design task. In *Design creativity 2010*. Springer, 257–264.

[68] Winger Sei-Wo Tseng. 2018. Can visual ambiguity facilitate design ideation? *International Journal of Technology and Design Education* 28, 2 (2018), 523–551.

[69] Barbara Tversky and Juliet Y Chou. 2011. Creativity: depth and breadth. In *Design creativity 2010*. Springer, 209–214.

[70] 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.

[71] Wen-Fan Wang, Chien-Ting Lu, Nil Ponsa i Campanyà, Bing-Yu Chen, and Mike Y Chen. 2025. AIdeation: Designing a Human-AI Collaborative Ideation System for Concept Designers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–28.

[72] Zhijie Wang, Yuheng Huang, Da Song, Lei Ma, and Tianyi Zhang. 2024. PromptCharm: Text-to-Image Generation through Multi-modal Prompting and Refinement. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–21.

[73] Zijie J Wang, Evan Montoya, David Munechika, Haoyang Yang, Benjamin Hoover, and Duen Horng Chau. 2022. Diffusiondb: A large-scale prompt gallery dataset for text-to-image generative models. *arXiv preprint arXiv:2210.14896* (2022).

[74] Justin D Weisz, Jessica He, Michael Muller, Gabriela Hoefer, Rachel Miles, and Werner Geyer. 2024. Design Principles for Generative AI Applications. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–22.

[75] Ou Wu, Weiming Hu, and Lei Shi. 2013. Measuring the visual complexities of web pages. *ACM Transactions on the Web (TWEB)* 7, 1 (2013), 1–34.

[76] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. 2022. Ai chains: Transparent and controllable human-ai interaction by chaining large language model prompts. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–22.

[77] David C Wynn, Khadidja Grebici, and P John Clarkson. 2011. Modelling the evolution of uncertainty levels during design. *International Journal on Interactive Design and Manufacturing (IJIDeM)* 5 (2011), 187–202.

[78] Suhyeon Yoo, Khai N Truong, and Young-Ho Kim. 2025. ELMI: Interactive and Intelligent Sign Language Translation of Lyrics for Song Signing. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–21.

[79] JD Zamfirescu-Pereira, Eunice Jun, Michael Terry, Qian Yang, and Björn Hartmann. 2025. Beyond code generation: Llm-supported exploration of the program design space. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–17.

[80] 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.

[81] Hongbo Zhang, Pei Chen, Xuelong Xie, Chaoyi Lin, Lianyan Liu, Zhuoshu Li, Weitao You, and Lingyun Sun. 2024. Protodreamer: A mixed-prototype tool combining physical model and generative AI to support conceptual design. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–18.

[82] Siyan Zhao, Mingyi Hong, Yang Liu, Devamanyu Hazarika, and Kaixiang Lin. 2025. Do llms recognize your preferences? evaluating personalized preference following in llms. *arXiv preprint arXiv:2502.09597* (2025).

[83] Qinyi Zhou, Jie Deng, Yu Liu, Yun Wang, Yan Xia, Yang Ou, Zhicong Lu, Sai Ma, Scarlett Li, and Yingqing Xu. 2025. ProductMeta: An Interactive System for Metaphorical Product Design Ideation with Multimodal Large Language Models. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–24.