DesignMinds: Enhancing Video-Based Design Ideation with a Vision-Language Model and a Context-Injected Large Language Model

He, Tianhao; Stanković, Andrija; Niforatos, Evangelos; Kortuem, Gerd

DOI
10.1145/3719160.3736633

Licence
CC BY

Publication date
2025

Document Version
Final published version

Published in
CUI '25: Proceedings of the 7th ACM Conference on Conversational User Interfaces

Citation (APA)
He, T., Stanković, A., Niforatos, E., & Kortuem, G. (2025). DesignMinds: Enhancing Video-Based Design Ideation with a Vision-Language Model and a Context-Injected Large Language Model. In J. Sin, E. Law, J. Wallace, C. Munteanu, & D. Korre (Eds.), *CUI '25: Proceedings of the 7th ACM Conference on Conversational User Interfaces* Article 55 ACM. https://doi.org/10.1145/3719160.3736633

DesignMinds: Enhancing Video-Based Design Ideation with a Vision-Language Model and a Context-Injected Large Language Model

Tianhao He
Delft University of Technology
Design at Scale Lab (D@S Lab)
Delft, Zuid-Holland, Netherlands
t.he-1@tudelft.nl

Andrija Stanković
Independent Industrial Design Engineer
Zagreb, Croatia
andrija.stkvc@gmail.com

Evangelos Niforatos
Delft University of Technology
Faculty of Industrial Design Engineering
Delft, Zuid-Holland, Netherlands
e.niforatos@tudelft.nl

Gerd Kortuem
Delft University of Technology
Faculty of Industrial Design Engineering
Delft, Zuid-Holland, Netherlands
g.w.kortuem@tudelft.nl

Figure 1: A designer in the experimental group is interacting with *DesignMinds*.

Abstract

Ideation is a critical component of video-based design (VBD), where videos serve as the primary medium for design exploration and inspiration. The emergence of generative AI offers considerable potential to enhance this process by streamlining video analysis and facilitating idea generation. In this paper, we present *DesignMinds*, a prototype that integrates a state-of-the-art Vision-Language Model (VLM) with a context-enhanced Large Language Model (LLM) to support ideation in VBD. To evaluate *DesignMinds*, we conducted a between-subject study with 35 design practitioners, comparing its performance to a baseline condition. Our results demonstrate that *DesignMinds* significantly enhances the flexibility and originality of ideation, while also increasing task engagement. Importantly, the introduction of this technology did not negatively impact user experience, technology acceptance, or usability.

CUI '25, Waterloo, ON, Canada

ACM ISBN 979-8-4007-1527-3/25/07
https://doi.org/10.1145/3719160.3736633

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**; • **Computing methodologies → Planning for deterministic actions**.

Keywords

Design Ideation, Generative AI, Video-based Design, Large Language Model, Vision Language Model, Eye-tracking, Designer-AI Collaboration

ACM Reference Format:
Tianhao He, Andrija Stanković, Evangelos Niforatos, and Gerd Kortuem. 2025. DesignMinds: Enhancing Video-Based Design Ideation with a Vision-Language Model and a Context-Injected Large Language Model. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25), July 08–10, 2025, Waterloo, ON, Canada.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3719160.3736633

1 Introduction

Idea generation is the cornerstone of innovation and serves as the foundation for new designs [14, 51]. Video-Based Design (VBD) enables designers to utilize video content as a key tool for generating knowledge, inspiring new ideas, and identifying potential challenges [26, 82, 84, 90]. The ideation of VBD plays a crucial role in brainstorming to produce a wide range of ideas, which are then filtered and refined to develop optimal solutions [20, 52, 53]. However, generating novel design ideas from videos is challenging for a large group of practitioners. It requires not only a significant investment of time and effort but also extensive design experience to generate a substantial number of related ideas for practice [20]. Consolidating design problems and generating feasible solutions from videos using traditional VBD methods typically requires extensive video review and the application of professional divergent thinking [90]. This process is often labor-intensive and heavily dependent on the practitioner's design experience and knowledge, which can be particularly challenging for novice designers with limited expertise and resources [96]. Additionally, previous research indicated that advanced video tools can potentially enhance the design work with videos to improve the quality outcomes, and to facilitate interactions [92].

With the recent surge in Generative AI (GenAI), technologies such as the Large Language Model (LLM) GPT-4 [58] demonstrate significant potential to enhance creative tasks across various design domains. A base LLM model can generate ideas across diverse scopes. Its capabilities can be further refined by incorporating contextual material through a process known as Retrieval-Augmented Generation (RAG) to make it adaptable in current circumstances [46]. Additionally, Vision-Language Models (VLMs) possess the ability to interpret videos with high detail, reducing the need for extensive human effort [8]. These advancements have the potential to assist designers in overcoming challenges associated with generating efficient and effective ideas, particularly when faced with prolonged video viewing and limited design experience [43, 66]. As such, this paper explores an approach that combines a customized VLM and LLM (*DesignMinds*) to enhance the "watch-summarize-ideate" process in VBD tasks through designer-AI co-ideation. We then evaluate the quality of the generated ideas, cognitive processes, user experience (UX) and technology acceptance and use from VBD ideation. Our work makes the following contributions:

- We introduce a novel GenAI-powered chatbot that features video understanding and design-context-based idea recommendations to enhance the ideation capabilities of new VBD practitioners.
- We investigate the impact of our prototype in terms of ideation quality, cognitive processing during ideation, and subsequent UX and technology acceptance.
- Ultimately, we propose a potential tool (*DesignMinds*) involving the use of a customized VLM and LLM to scale up the VBD ideation process for new designers.

Our findings indicate that *DesignMinds* improves the flexibility and originality of design ideas and boost design task engagement. The adoption of this technology also did not adversely affect the established patterns of UX, technology acceptance and usability.

2 Background

2.1 Ideation in Design

In the design process, ideation is a key aspect of experience that influences both the initiation and progression in the early stage of creative activities. Eckert and Stacey articulated that ideation is not merely a catalyst for creativity but also a critical component in developing design ideas [25]. They claimed that ideation in design provides a contextual framework that enables designers to effectively communicate and position their work. It sparks design creativity, offering new perspectives and triggering the generation of original ideas [25]. Similarly, Setchi and Bouchard define ideation as a multifaceted phenomenon where designers absorb and reinterpret existing ideas, forms, and concepts [73]. This process is influenced by designers' individual experiences, cultural backgrounds, and personal interests and serves as a guiding principle for creativity. The subjectivity of ideation accelerates designers to explore a broader array of possibilities. Gonçalves et al. extended the understanding of ideation into later stages, asserting that designers maintain a limited range of external stimuli preferences. Both design students and professionals often favor visual stimuli such as images, objects, and video sources to encourage creativity [27].

However, relying on specific stimuli and designers' own knowledge may cause the risk of design fixation [38]. This phenomenon occurs when designers over-rely on specific knowledge directly associated with a problem or themselves during ideation, eventually inhibiting the design outcome [50, 91]. Viswanathan and Linsey claimed that the problem of fixation is pervasive and varies inversely with the level of design expertise. They suggested that it is especially prominent among novice designers, who tend to rely heavily on their predominant knowledge during ideation [85]. In addition, novice designers often struggle to analyze problems comprehensively and have difficulty seeking helpful information during ideation [18, 24]. This phenomenon often leads to failures in framing problems and directing the search for solutions, ultimately diminishing the design outcome. Gonçalves pointed out that the lack of reflection in ideation could be addressed by developing computational tools to help designers efficiently find relevant stimuli. Such tools could assist inexperienced designers in exploring ideas that are semantically distant from the problem domain and expand space for ideation [27]. Similarly, the study by Dazkir et al. showed that while self-selected contexts in designers led to greater interest in the topic, they often failed to develop effective design solutions. This indicates that, although some autonomy is beneficial for developing design ideas, many inexperienced designers still need external intervention in the early stages to aid in ideation [21]. As such, designers, especially those with limited experience, often need additional help and guidance from outside sources to enhance ideation.

2.2 Videos for Design Ideation

The use of video as a central tool for ideation, known as VBD, involves capturing information and analyzing solutions in design process. This technique is particularly prevalent in fields such as user experience (UX) design, interaction design, and ethnographic research [90]. By recording user interactions with products or environments, videos provide a dynamic and context-rich data source for designers. Design videotapes are informative for practitioners to deepen context understandings and generate follow-up interventions [89, 90]. Unlike textual descriptions from interviews or surveys, videos preserve temporal and sequential nuances and allow designers to revisit specific moments repeatedly for deeper analysis [82]. Designers at Apple Inc., for example, utilized videos to envision new user interfaces (UIs) for their future computers [84]. They utilized videos to benchmark new UIs and study user behavioral reactions through videotapes. Similarly, Tatar from PARC explored learning from repeated video observations of user behavior through stationary camera recordings and aimed to minimize erroneous assumptions in software development [82]. Tatar also emphasized the important role of using videos for ideation to pinpoint design solutions. Ylirisku and Buur further conceptualized this practice and highlighted that using videos for design ideation is instrumental for practitioners. Videos are an effective tool for learning from target users' daily experiences and augment designers generate an abundance of ideas for design artifacts [89]. Moreover, designers can ideate from the "thick descriptions" that videos capture about users' movements, interactions, and emotional transitions, which help in constructing design narratives and encapsulating individual thoughts.

Other media and methods have been explored for supporting design ideation, yet each shows limitations compared to videos. Visual imagery, such as photographs or example collections, is commonly used in graphic design practice to inspire creativity [56, 78]. However, empirical studies show that while imagery exposure may aid in personal development and idea communication, its impact on creative quality and novelty is limited, and risks inducing design fixation [38, 45]. Similarly, AI-generated imagery can potentially open fresh avenues for artistic expression, but tends to produce outcomes that are often disconnected from the designer's original intent and require considerable reinterpretation and adaptation [12]. Additionally, research by Delle Monache et al. has explored how sonic stimuli can stimulate ideation by focusing on emotional and semantic sound attributes [22]. While their methods promote group iteration and diverse thinking, they rely heavily on subjective interpretation through verbal and vocal expressions. This reliance limits designers' ability to capture temporal and environmental context in a broader context [90]. Beyond sound-based methods, text-driven composition often relies on abstract representations rather than embodied experiences and thus lacks the immediacy and contextual dynamics crucial for user-centered design [17, 19]. This reliance tends to reinforce familiar thinking patterns [38] and limits exploration of the broader solution space. In contrast, videos capture the temporal progression of real-world interactions, emotional expressions, and environmental dynamics [82, 89, 90] which helps designers to repeatedly reengage with authentic user behaviors.

While video-based design idea generation presents significant opportunities, videos often contain complex content and frequent events [89, 90]. The process of watching these videos can be labor-intensive and time-consuming. Videos with rich details and rapid sequences require from viewers substantial information processing effort to analyze perceived information. As a result, designers may suffer risks of diminishing decision-making capability and result in a decline in ideation effectiveness [9, 60]. Therefore, it is essential to develop strategies to mitigate fatigue and reduce the information processing effort for designers who use videos for inspiration, while ensuring that they retain the valuable information presented in videos.

2.3 GenAI for Design Ideation

Recent advancements in GenAI are driving significant changes across multiple disciplines. Large Language Models (LLMs), such as GPT-4 [58], have shown remarkable capabilities in assisting creative tasks for design purposes [94]. Xu et al. proposed an LLM-augmented framework that uses LLM prompts to generate unified cognition for practitioners and optimize the creative design process in a professional product design [87]. Another group of researchers proposed Jamplate, a protocol that leverages formatted prompts in LLMs to guide novice designers in real-time. This approach enhances their critical thinking and improves idea generation more effectively [88]. Makatura et al. explored the use of GPT-4 to generate textual design language and spatial coordinates for product design and adaptation in industry [55]. They highlighted that GPT-4's reasoning capabilities offer significant value in novel design domains. When designers are inexperienced with a particular domain or working on a novel problem, GPT-4 can synthesize information from related areas to provide suitable advice.

Beyond purely text-based ideation, researchers have started to explore multimodal models that combine language and vision. VLMs, which extend LLMs by incorporating visual input capabilities, offer new possibilities for grounding ideation in visual contexts [47]. For example, Zhou et al. introduced NavGPT, an LLM-based navigation agent that uses visual cues detected by a VLM to provide indoor navigation suggestions [95]. They demonstrated that their system can generate high-level navigational suggestions from automatic observations and moving histories. Moreover, Picard et al. explored the use of GPT-4V(ison) [58], a version of GPT-4 with vision-language capabilities, in product design. They investigated its application in design tasks, such as analyzing handwritten sketches and providing follow-up suggestions for material selection, drawing analysis, and spatial optimization. Their findings demonstrated that this LVM model can handle complex design idea generation with proficiency [63].

Recent studies show that integrating VLMs and LLMs is a more complete solution for high-level creative tasks [36, 59, 62]. VLMs are better at extracting salient information and contextual hints from visual data such as images or video frames [47], whereas LLMs possess superior ability in abstract reasoning, conceptual generation,

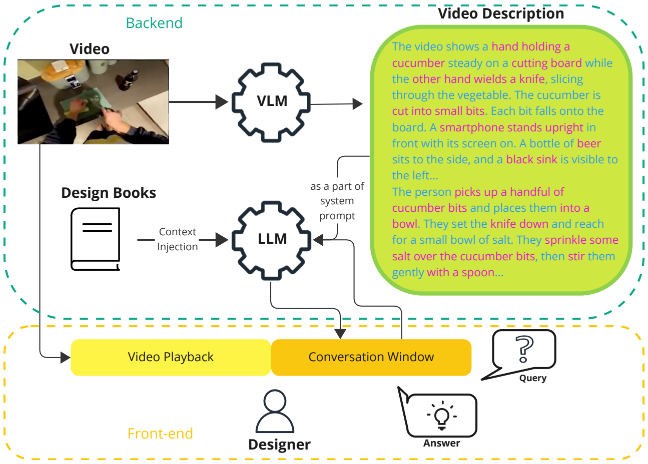


Figure 2: *DesignMinds* consists of two primary components: the backend and the front-end. The backend includes a VLM and a LLM integrated with a design knowledge repository. The front-end features a video playback region alongside a conversational window. The videos are first processed to extract key terms (highlighted in pink in video description) and are then connected into a comprehensive description (blue in video description) using in-built language linking functions. These complete descriptions are then passed to the LLM, along with a knowledge repository enriched by selected design books from a committee vote. Designers can then use the features in front-end to watch the video playback to enhance trust and grounding for the design context, and engage in ideation through conversations in the conversational window.

and storytelling generation [11, 57]. However, each model class in itself is restricted: VLMs are challenged to generate extended abstract ideation from non-visual content, and LLMs without visual anchoring are capable of generating bland or context-mismatched responses [58]. The merging of LLMs and VLMs presents an opportunity to enhance design, prompting us to explore this integration in the refined field of VBD. Motivated by these insights, we explore whether combining LLMs and VLMs can benefit VBD practitioners in generating design ideas. To investigate, **we prototyped *DesignMinds* that integrates a state-of-the-art (SOTA) VLM and LLM model with a context-injection technique. We conducted a user study involving two video-based design tasks to assess the impact on design ideation, focusing on ideation quality, cognitive processes, UX, and technology acceptance**.

3 Our *DesignMinds* Prototype

The development of our prototype followed the natural process of the idealization of VBD, consisting of two main parts: video comprehension and idea reflection and refinement [89]. As shown in Fig. 2, *DesignMinds* consists of two primary components: the backend and the front-end. The backend includes a VLM and a LLM integrated with a design knowledge repository for reference. The front-end features corresponding a video playback region alongside a conversational window. We adopted `blip2-opt-6.7b`[1], a SOTA VLM, to interpret videos into textual descriptions. When processing a video, the VLM first extracts perceived objects from the video and utilizes built-in language connection functions to generate comprehensive textual descriptions of the entire video. These complete video descriptions then were processed by an LLM through GPT-4 API (`gpt-4-0125-preview`)[2]. To generate more design-grounded suggestions, we implemented a RAG function using a text embedding model `text-embedding-ada-002`[3] on a framework of LlamaIndex[4] as our *DesignMinds*'s professional knowledge repository for conversations. To ensure that the knowledge repository provided designer-relevant information for our LLM, we conducted a discussion on VBD literature within an independent community of designers (N = 30). This discussion led to a vote that selected six authoritative books (1,966 pages total) with high-level methodological rigor and practical design cases for the VBD training. We then utilized the RAG function and tokenize the selected design books to feed into the knowledge repository of the LLM. We then built our front-end interface using Gradio[5] as illustrated in Fig. 3. The interface includes a video player and an chatbot conversation window. To test performance and enhance convenience for test users in the later study, we allocated the right portion of the screen to included a text box where users could record their ideas and inspirations. This setup allows users to review and revisit the design context using the video player, generate additional insights and ideas through the chatbot, and record their comprehensive thoughts in the text box for later use.

4 Study

We evaluate how our proposed *DesignMinds* influences ideation in VBD tasks with a between-subject study design. Specifically, we examine whether and how the tool influences designers' effectiveness and ability to generate ideas from video content. Our assessment is structured around three key perspectives: the quality of ideas generated by designers, the cognitive processes they undergo during the ideation tasks, and their overall user experience and acceptance of the new prototype. Additionally, we analyze how designers interact with *DesignMinds* from their conversation logs to better understand the ideation process. We also explore *DesignMinds*' potential cognitive effects, perceived usefulness and likelihood of adoption by designers. Finally, we investigate areas for improvement and suggest ways to enhance *DesignMinds*' usability and other concerns. Our study addresses the following Research Questions (RQs):

RQ1 *How does DesignMinds influence the quality of ideas generated in the VBD process?*
Divergent thinking, a concept introduced by Guilford [30, 31] acts as a foundational idea in creativity research. In design, it also allows VBD practitioners to identify innovative solutions based on available resources [1, 6, 30, 90]. Building on this foundation, we investigate how our *DesignMinds* impacts

[1]https://huggingface.co/Salesforce/blip2-opt-6.7b (last accessed: May 25, 2025).
[2]https://platform.openai.com/docs/models/gpt-4-turbo-and-gpt-4 (last accessed: May 25, 2025).
[3]https://platform.openai.com/docs/guides/embeddings (last accessed: May 25, 2025).
[4]https://docs.llamaindex.ai/en/stable/ (last accessed: May 25, 2025).
[5]https://www.gradio.app/ (last accessed: May 25, 2025)

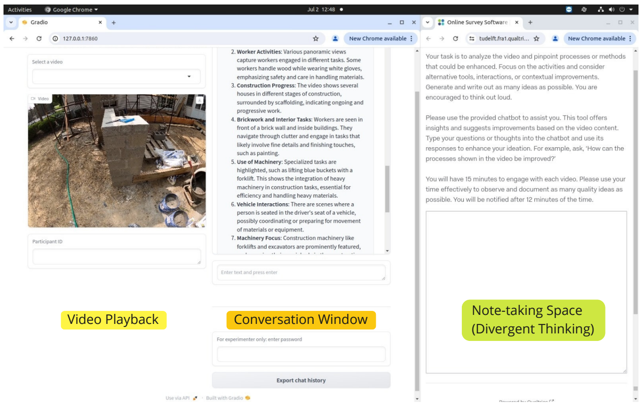


Figure 3: The interface of *DesignMinds* primarily features a video player on the left and an LLM conversation window in the center. To facilitate organized ideation recording in the later study, we additionally included a note-taking space bellow a description of VBD tasks for recording participants' divergent thinking during the study tasks (see Supplementary Text 1) for detailed text. When designers use *DesignMinds*, the system initially performs a background pre-analysis of the video content on the left, and transitions video content to the chat interface in the center. Designers subsequently interact via chatting and generate inspiration as Divergent Thinking notes on the right.

the outcomes of divergent thinking by asking participants in two conditions (experimental group and control group) to generate creative ideas during the task. We hypothesize that **designers with AI co-ideation will exhibit higher Divergent Thinking scores compared to ideation without AI**.

RQ2 *How does DesignMinds influence the way designers practice ideation in VBD?*
Examining user behaviors is another critical aspect of evaluating the VBD ideation process, in addition to assessing the final deliverables. The behaviors exhibited during tasks reflect participants' approaches to completing the assigned tasks [4, 5, 32]. We record their eye movements to evaluate the level of engagement and cognitive load experienced by designers in both conditions. Additionally, we conduct an in-depth analysis of the chat log history from the experimental group to understand how participants interacted with *DesignMinds*. We hypothesize that **designers will experience greater engagement and, consequently, a slightly higher cognitive load in the AI-prototype-assisted condition**.

RQ3 *What impact does DesignMinds have on the User Experience (UX) and Technology Acceptance and Use in the VBD ideation process?*
The introduction of new technologies or tools to a traditional methodology can sometimes cause discomfort and decreases in UX [80]. Understanding and evaluating technology acceptance and use also provides insights into how well users adapt to new technology, which may potentially impact the original practice. We further compare the UX and the level of acceptance and use of technology between our prototype condition and the control condition during the VBD ideation process. We hypothesize that **the newly introduced prototype will not have additional negative influence on UX and technology acceptance and use compared to traditional practices**.

4.1 Participants

Table 1: The demographics of participants' design experience, including possible responses and their values, are presented as answer frequencies (f), followed by the corresponding percentages (%).

Variable	Answer	f	%
Current design educational level	Bachelor	12	34.29%
	Master	22	62.86%
	PhD (ongoing)	1	2.86%
Experience of designing with videos (VBD)	Definitely not	15	42.86%
	Probably not	7	20.00%
	Might or might not	9	25.71%
	Probably yes	2	5.71%
	Definitely yes	2	5.71%
Experience of practicing design divergent thinking (ideation)	Definitely not	2	42.86%
	Probably not	2	5.71%
	Might or might not	9	25.71%
	Probably yes	17	48.57%
	Definitely yes	5	14.29%
Proficiency in using chatbot	Never used before	2	5.71%
	Beginner	7	20.00%
	Intermediate	13	37.14%
	Expert	13	37.14%

We enlisted 35 design graduates (17 females and 18 males) from the design faculty at our university, following approval from the ethics board and confirming that none had any cognitive impairments. The participants, who are either university students (BSc & MSc) or PhD candidates, had an average age of 25.4 years (SD = 2.31) and an average of 2.4 years of design experience (SD = 1.14). Table 1 presents the demographics of participants involved in the study, including their educational levels, self-assessed familiarity with VBD experience and ideation, as well as their proficiency in using chatbots such as ChatGPT. In addition, participants with visual acuity below 20/20 were instructed to wear contact lenses before participating. All participants were fully informed and provided consent before the experiment began.

4.2 Apparatus

In our experiment, we evaluated our system in an office setting with consistent lighting. The system was set up to operate as localhost on a desktop computer within the lab environment. Fig. 1 illustrates the lab setup where participants engaged with the system. Alongside standard office equipment such as a keyboard, mouse, and speaker, participants were asked to wear eye-tracking glasses (Pupil Labs[6]). These glasses were connected to a phone record eye movement data. Additionally, we placed four AprilTags[7] on each corner of the monitor (see Fig. 1) to allow eye-tracking glasses to detect the screen's edges and define areas of interest (AOI).

4.3 Measures

4.3.1 *Subjective Measures.*

- **Evaluation of Divergent Thinking (RQ1)**: we employ an established protocol of divergent thinking [30, 31] and assess it through the following three dimensions:
  - **Fluency**: the quantity of comprehensive ideas with sufficiently details generated [6].
  - **Flexibility**: the range of different domains and subdomains covered by the ideas [1, 67, 70].
  - **Originality**: the statistical infrequency of ideas [68, 70].
- **Chat Log history (RQ2)**: the intermediate conversation history made by participants in the experimental group with AI co-ideation using the chatbot of *DesignMinds*.
- **Unified theory of acceptance and use of technology (UTAUT) (RQ3)**: a widely recognized model for assessing how users accept and adopt information technology considers the perceived likelihood of adoption [3, 83, 86].
- **User Experience Questionnaire (UEQ) (RQ3)**[8]: a questionnaire designed to measure UX in interactive products uses a benchmarking method that organizes raw UEQ scores into categories such as efficiency, perspicuity, dependability, originality, and stimulation [71, 72].

4.3.2 *Objective Measures (RQ2).*

- **Pupil dilation**: variation in pupil diameter as an indicator of cognitive load and decision-making engagement [7, 23]
- **Fixation**: stationary periods of the eyes over a particular point, associated with focused attention and information processing [35]
- **Blink rate**: the frequency of blinks, reflecting emotional interest and cognitive engagement [54, 77]
- **Saccade**: rapid, synchronized movements of the eyes between fixation points, with saccade velocity reflecting visual search and engagement [65]

4.4 Procedure

Participants were first assigned to either the control or experimental group using de-identified IDs. They were then individually invited to the lab and took their designated positions in front of the monitor (see Fig. 1). They were introduced to the study apparatus, including the user interface (shown in Fig. 3) relevant to their assigned group, how to wear the eye-tracking glasses, and briefed on the study procedure. After this introduction, participants were asked to complete a consent form and provide demographic information, including their experience with design ideation from videos, general design experience, and familiarity with using chatbots. Then

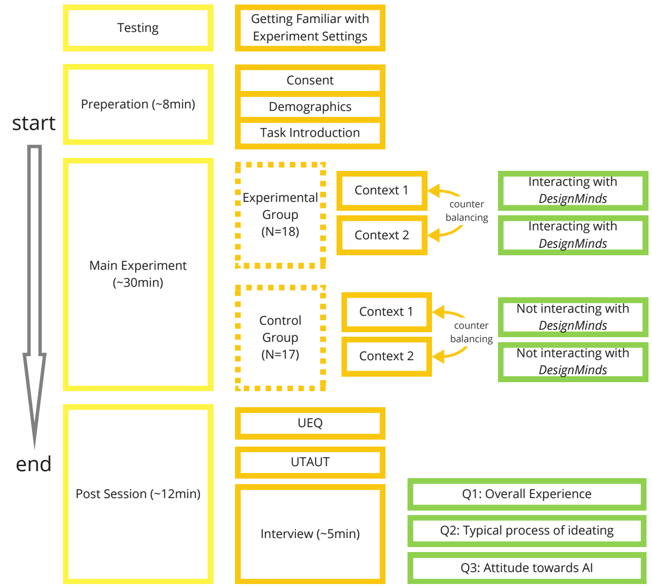


Figure 4: During the study, participants were initially asked to familiarize themselves with both the environment and *DesignMinds* (Testing). They received instructions on the components of the prototype and how to interact with it. After completing consent and demographic forms, participants were provided with preparation instructions for tasks. First, participants were randomly divided into two groups: the experimental group, which interacted with the chatbot *DesignMinds*, and the control group, where participants continued their usual practice for design inspiration. Each participant group was assigned two tasks with different design contexts, presented in a counterbalanced order. After finished with main experiment, participants were asked to complete the UEQ and UTAUT questionnaires. They were then interviewed on three topics: overall experience, typical ideation process, and their attitudes towards AI.

they received instructions (see Supplementary Text 1) on the tasks they were required to complete. Following the preparatory phase, participants in each group were shown two video tasks depicting contexts of cooking and construction, with the order of presentation counterbalanced. These videos were sourced from Ego4D[9], a large-scale video dataset frequently employed for benchmark and HCI research [28]. Considering the total length of the study, each video was approximately 3 minutes in duration. In the experimental group, participants were instructed to use the defined UI shown in Fig. 3 to watch video playback, interact with the chatbot, and make notes in the designated note-taking space to record divergent thinking. In contrast, for the control group, the chatbot was hidden, and participants were asked to proceed with design ideation on the note-taking space from the videos as they normally would. Participants were notified at the 12-minute mark of each task that they had 3 minutes remaining. The process was repeated for both videos. Upon completion, participants were then asked to evaluate

[6]https://pupil-labs.com/products/neon/ (last accessed: May 25, 2025).
[7]https://april.eecs.umich.edu/software/apriltag (last accessed: May 25, 2025).
[8]https://www.ueq-online.org/ (last accessed: May 25, 2025).
[9]https://ego4d-data.org/ (last accessed: May 25, 2025).

their experience using UEQ and UTAUT questionnaires. Following the questionnaires, a 5-minute interview asked participants insights toward three topics: their overall experience during the two video tasks, their performance in the ideation process, and attitudes toward AI after having the experiment.

5 Results

5.1 Divergent Thinking Analysis (RQ1)

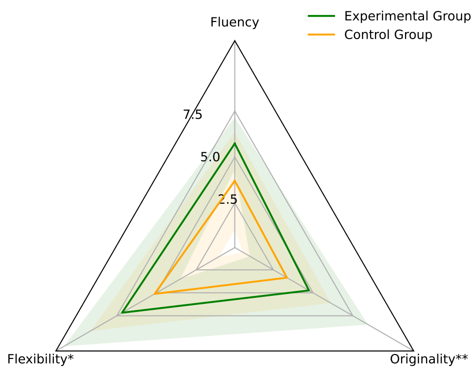


Figure 5: Radar chart depicting the evaluation scores of design thinking across raters for the experimental and control groups. Errors are indicated by shaded regions. Attributes marked with asterisks (* or **) represent significant differences. * denotes 0.01 < p < .05, and ** denotes p < .001.

To address the quality of ideas generated in the VBD process as proposed in RQ1, we collected the divergent thinking texts from both groups. We then recruited three independent raters to evaluate the ideation results based on fluency, flexibility, and originality, using a predetermined set of criteria (See supplementary text 2) [30]. We then performed a quantitative analysis of the rating scores for both the experimental and control groups. As shown in Fig. 5, we observed a significant main effect on the average ratings for flexibility and originality (independent t-test $t(33)t = 2.304$, $p = .014$; $t(33)t = 4.674$, $p < .001$). **The average scores for both flexibility** (7.17 ± 3.511 points) **and originality** (4.74 ± 1.018 points) **in the experimental group were significantly higher than those in the control group** (flexibility: 5.12 ± 1.074 points; originality: 3.35 ± 0.583 points). However, there was no significant main effect on the rating for fluency between the two groups (independent t-test $t(33)t = 1.885$, $p = .068$). Additionally, Krippendorff's Alpha was calculated to assess the internal consistency of the three raters' judgments on the categories of divergent thinking. We observed a moderate agreement among the raters, with an average Krippendorff's Alpha of $\alpha = .702$ (95% CI, .245 to 1), p < .001.

5.2 Design Ideation Process (RQ2)

5.2.1 *Eye-tracking measures.*

We first analyzed the eye-tracking results from both groups. As shown in Fig. 6c, a significant main effect was observed in the average pupil dilation between the experimental and control groups (independent t-test $t(33)t = 2.933$, $p = .021$). The dashed line in the subplot represents 0 millimeters which indicates there was no change from participant's baseline pupil diameter during non-tasked time. Compared to the baseline, **participants in the experimental group exhibited an average dilation of 0.15 mm more than those in the control group during the ideation task** (experimental $std = 0.206$; control $std = 0.152$). We then examined the gaze fixation rate per minute and the average fixation duration across the two groups. As shown in Fig. 6b, no significant main effect (independent t-test $t(33)t = 0.795$, $p = .986$) was observed in the average fixation rate (see subplot (a)). Interestingly, **participants in the experimental group exhibited an average fixation duration that was significantly 120.31 milliseconds** ($std = 135.053$) **longer than that of the control group within the AOI** ($std = 193.366$; independent t-test $t(33)t = 1.567$, $p = .039$). Additionally, as shown in subplot (a) of Fig. 6c, we observed a significant main effect in the average blink rate (independent t-test $t(33)t = 0.557$, $p = .004$). **Participants in the experimental group on average blinked 5.23** (experimental $std = 4.459$; control $std = 5.400$) **less times per minute than those in the control group**. However, no significant difference was found in the average blink duration between the two groups (independent t-test $t(33)t = 0.226$, $p = .340$). No significant main effect was observed in the average saccade rate between the two groups shown in Fig. 6d(independent t-test $t(33)t = 0.252$, $p = .249$). However, there was a significant increase in saccade velocity in the experimental group compared to the control group (independent t-test $t(33)t = 3.171$, $p < .001$). On average, **participants in the experimental group performed 662.45 pixels per second faster saccades than those in the control group within the AOI** (experimental $std = 477.332$; control $std = 351.452$).

5.2.2 *Chat log analysis.*

In addition to eye-tracking measurements, we conducted an in-depth analysis of the conversation logs from the experimental group. We utilized both qualitative and quantitative methods to better understand what occurred during the augmented design ideation processes with *DesignMinds*. We categorized the questions that participants asked as follows:

(a) Questions about design opportunities (N=16): The majority of questions posed by participants (P1-3, P5, P7-17, and P19) focused on suggestions or ideas for improving the processes depicted in the videos. These inquiries typically emerged after participants had gained an understanding of the video's content and identified key areas of interest for potential design opportunities. For instance, some designers, such as P2 and P8, sought initial inspiration to begin their designs by asking, "How can the processes shown in the video be improved?" (P2) and "What can be improved?" (P8). Others (P3, P9, P12, and P19) aimed to build upon existing ideas and leveraged the LLM to further extend their concepts. These participants asked questions such as, "What do you suggest to avoid using hands directly when handling food during cooking?" (P9), "Can you recommend structures that allow a construction worker to lift heavy objects without carrying them?" (P12), and "What are the consequences of not using fitted kitchen tools for the task?" (P19).

(b) General video content understanding (N=13): Many participants (P2-4, P6-8, P10, P11, P13, P15, P16, P17 and P19) utilized the video comprehension capabilities of *DesignMinds* to gain a comprehensive understanding of the content presented in the videos. Participants frequently inquired about the events occurring in the video or sought clarification on specific actions or objects they found unclear. Some participants employed a *DesignMinds*-first strategy, initiating their ideation processes by querying the LLM about the video's content. For example, common inquiries included, "What is this video about?" (P2), "List the steps of the activities." (P6), "What dish is he making?" (P11), and "Can you tell me what's happening in the video?" (P15). Others used *DesignMinds* to validate their observations, asking questions such as, "Are they cutting the edge in a straighter line?" (P10) and "This video was about how to cut an avocado, right?" (P17). Additionally, a subset of participants posed higher-level, reflective questions about the video's content, such as P19, who asked, "What is the goal of what they are doing during the construction work?"

(c) Understanding and Ideation from Specific Scene Settings (N=10): A subset of participants (P3, P6, P7, P9, P12-14, P16, P17 and P19) sought to utilize *DesignMinds* to gain a deeper understanding of specific scene settings depicted in the videos. Unlike the broader inquiries in category (b), these participants focused on more narrowly defined actions within a given context. For example, when viewing a scene where an individual attempts to retrieve food from a sealed jar, P6 asked the LLM, "What are some ways to lock a jar automatically?" Similarly, P9 used the prototype as a tool for identifying specific items, asking, "What is the tool called that slices cheese in this video?" P14 inquired about strategies for organizing kitchen utensils, asking, "Can you combine the relocation ideas for kitchen tools?" In the context of construction, P16 sought detailed advice by asking, "How can I make sure that the men operate heavy machinery safely?" while P17 questioned, "Which is more efficient: adding an extra step in the process or using two different tools?"

(d) Combining Impressions with Opinion-Based Queries (N=4): Some participants (P6, P11, P16, P19) went a step further by integrating their own impressions with their questions and aksed for pinion-based suggestions. For instance, P11, while observing a scene involving three workers in a construction setting, asked, "Don't you think the space is crowded for 3 people?" The participant here showcased a critical evaluation of the scene. Similarly, other participants framed their questions in a way that encouraged critical thinking. For example, P19 asked, "What happens if you don't use fitted kitchen tools for the job?"

Additionally, we conducted correlation tests to explore the relationship between traits from the chat logs during ideation and the quality of the final ideation outcomes, measured by three attributes: fluency, flexibility, and originality (see Fig. 5). We analyzed the conversation history and computed the average number of chat turns participants made with the prototype, the average number of words in each question asked and response generated, and the number of follow-up ideas generated for each participant in the experimental group. As shown in Table 2, Pearson product-moment correlation tests were conducted to measure the relationship between chat log variables and ideation quality. There was **a strong, positive correlation between the average number of words in each participant's question and the originality of the ideas ultimately generated**, which was statistically significant (ρ = .500, n = 18, p = .034). Similarly, **a strong and significantly positive correlation was found between the average number of words in each generated answer and the fluency** (ρ = .636, n = 18, p = .005), **flexibility** (ρ = .743, n = 18, p < .001), **and originality** (ρ = .652, n = 18, p = .003) of the ideation quality. In addition, **a strong and significantly positive correlation was also observed between the average number of ideas generated from the prototype and both the fluency and flexibility** of the ideation quality (ρ = .749, n = 18, p < .001; ρ = .782, n = 18, p < .001).

5.3 UX, Technology Acceptance and Use (RQ3)

To determine if the introduction of a new technology affected the ideation process from VBD, we analyzed self-reported data on participants' UX and technology acceptance from both the experimental and control groups, as shown in Fig. 7. We conducted one-way ANOVA and Kruskal-Wallis H tests for each attribute pair. The null hypothesis (H0) for these statistical tests assumed that there was no significant main effect between the two groups regarding attributes from UX and technology acceptance and use, meaning that the self-reported perceptions in both groups were the same. For the UEQ which measures UX (see Fig. 7a), the analysis revealed no significant difference in the attractiveness attribute between the experimental group that used *DesignMinds* and the control group as a baseline (ANOVA $F_{1,33}$ = .386, p = .538). Similarly, comparisons of the other UEQ attributes—perspicuity (ANOVA $F_{1,33}$ = 1.208, p = .332), efficiency (ANOVA $F_{1,33}$ = .008, p = .944), dependability (ANOVA $F_{1,33}$ = 0.200, p = .665), stimulation (ANOVA $F_{1,33}$ = 0.376, p = .553), and novelty (ANOVA $F_{1,33}$ = 1.639, p = .345)—between the experimental and control groups also retained the null hypothesis (H0). Thus, **all six UEQ attributes collected from the experimental group using *DesignMinds* measured UX has the same results as in the control group**.

Additionally, as shown in Fig. 7b, the non-parametric Kruskal-Wallis test revealed that the PE attribute (performance expectancy) from UTAUT failed to reject the null hypothesis ($x^2(1)$ = .003, p = .960), indicating no significant difference in performance expectancy between the groups. The mean rank scores were 17.92 for the experimental group, 18.09 for the control group. Similarly, the attributes of EE (effort expectancy) (ANOVA $F_{1,33}$ = 1.413, p = .081), ATT (attitude toward using technology) (ANOVA $F_{1,33}$ = .699, p = .287), ANX (anxiety) (ANOVA $F_{1,33}$ = .391, p = .442), and BI (behavioral intention) (ANOVA $F_{1,33}$ = .004, p = .938) also retained the null hypothesis. As such, **all attributes for measuring technology acceptance and use retained the null hypothesis between the two groups. These findings indicate that our experimental *DesignMinds* did not introduce any negative effects on UX or technology acceptance and use compared to the normal ideation process in VBD (control)**.

Table 2: Table of Pearson's correlation coefficients (ρ) and their p-values for four test variables from the analysis of the intermediate chat log and three ideation quality variables (see Fig. 5). Significant correlations are indicated by ** or * based on the p-values (see notes).

Chat Log Variable	Ideation Quality	Pearson's correlation coefficient	*P*-value
Avg. Nr. of Chat Turns	Fluency	0.261	0.296
	Flexibility	0.126	0.619
	Originality	-0.313	0.206
Avg. Nr. of Words in Each Question Asked	Fluency	0.218	0.385
	Flexibility	0.318	0.198
	Originality	.500*	0.034
Avg. Nr. of Words in Each Answer Generated	Fluency	.636**	0.005
	Flexibility	.743**	<.001
	Originality	.652**	0.003
Avg. Nr. of Ideas Generated	Fluency	.794**	<.001
	Flexibility	.782**	<.001
	Originality	0.398	0.102

Notes: Pearson's correlation test (two-tailed) is significant at $^{**}p < 0.01$ and $^{*}p < 0.05$.

6 Discussion

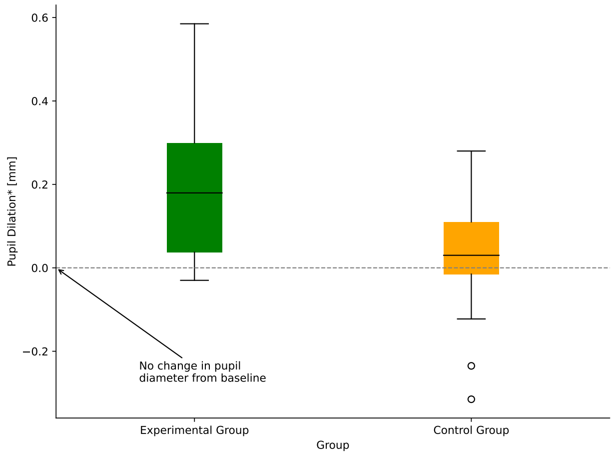
In this study, we conducted an A/B test to evaluate the impact of our *DesignMinds* on the ideation process for VBD. Participants were assigned two sub-tasks and asked to generate as many design ideas as possible related to the provided contexts. Our findings indicate that *DesignMinds* significantly enhanced participants' performance in terms of the flexibility and originality of their final ideation outputs compared to the baseline. Additionally, participants using *DesignMinds* demonstrated greater engagement in decision-making, as evidenced by eye-tracking data, and there was a strong positive correlation between the number of ideas and words generated with *DesignMinds* and the overall quality of their ideation. Furthermore, our findings suggest that the introduction of *DesignMinds* did not negatively impact user experience or technology acceptance.

6.1 Increased Flexibility and Originality in Divergent Thinking

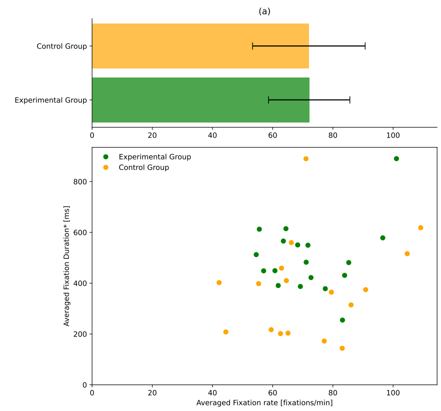
Divergent Thinking is a well-established method, supported by both theory and practice, in measuring creativity during ideation [67–69]. Because our prototype aims to support the generation of diverse and novel ideas through conversational assistance, Divergent Thinking offers a theoretically grounded lens to capture these creative outcomes. In this study, we adopted this approach to investigate how *DesignMinds* incorporating emerging technologies can enhance ideation within a design context involving videos. Our first research question (RQ1) explores the impact of *DesignMinds* on ideation outcomes. To address this, we collected Divergent Thinking data from our study and had three independent graders with an "internal consistency" check to evaluate the quality of ideation, following principles outlined in well-established literature [30]. At the outset, we reviewed how ideation is understood and measured in the literature. For instance, fluency is used to assess the productivity of ideation, while flexibility indicates diverse ideas across different conceptual categories. Originality is defined by the novelty or rarity of ideas within a given task [30]. Our results show that participants in the experimental group, supported by *DesignMinds*, received higher ratings in flexibility and originality compared to the control group. **This suggests that with *DesignMinds*' assistance, the ideation process generated more multifaceted and novel ideas** [68]. Specifically, the trait of flexibility could improve professional practitioners' understanding of tasks (e.g., the usability of an artifact) and decision-making in design projects (e.g., plans for improvement) [2]. Whereas originality, on the other hand, not only strongly correlates with innovation but also reflects the quality of authenticity and integrity of creative tasks [29]. Similarly, other studies concluded that ideation from industrial design tasks should consider three key aspects: "functional value", "aesthetic value" (e.g., visual form), and "originality value" [13]. Our study showed that the prototype notably improved outcomes in two of these aspects—flexibility and originality. As such, the use of *DesignMinds* enhanced the variety and novelty of ideas in creative VBD tasks.

6.2 Greater Engagement in Ideation and Positive Correlation Between Interaction History and Performance

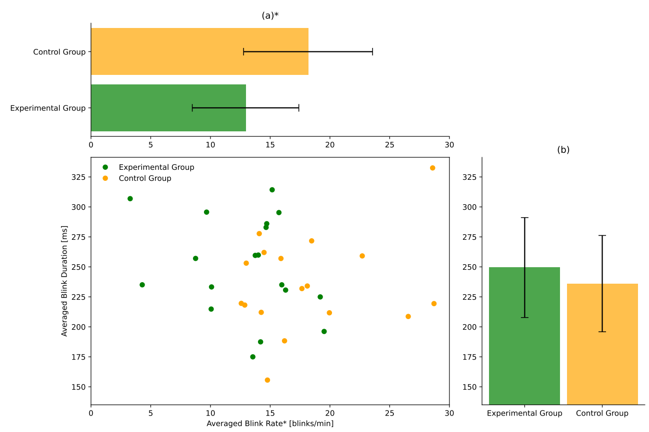
We found that the final outcome of ideation (i.e., Divergent Thinking) was partially enhanced by the prototype. To explore further, we sought to understand how our prototyped *DesignMinds* influenced the ideation processes in design tasks (RQ2). We began by measuring participants' eye movements during the tasks and observed an increased in pupil dilation in the experimental group compared to the control group. Previous studies have shown that dynamic changes in pupil dilation are associated with high-level cognitive processing [33]. Since the study was conducted in a stable lighting environment, the observed increase in pupil dilation indicates that participants voluntarily engaged in deeper, high-level decision-making prompted by the recommendations generated by



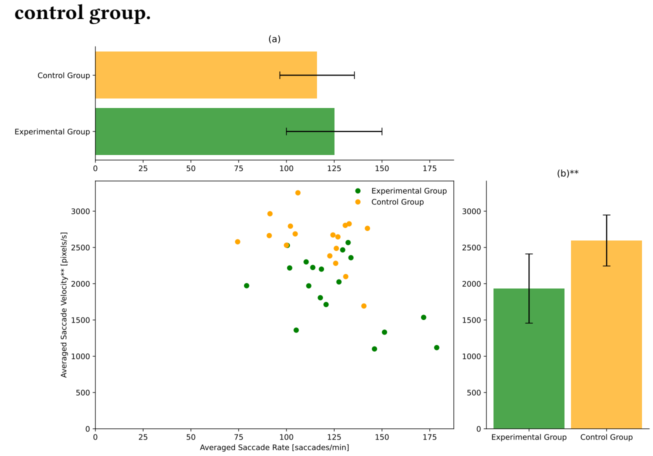
(a) Participants in the experimental group exhibited significantly greater pupil dilation compared to the control group. The dashed line at 0 millimeter on y axis represents no change in pupil diameter relative to the baseline, when participants were not engaged in ideation tasks.



(b) In subplot (b-a), no significant difference was observed in the averaged fixation rate between the groups. In subplot (b-b) indicated by an asterisks(*), participants in the experimental group exhibited a significantly higher fixation duration compared by control group.



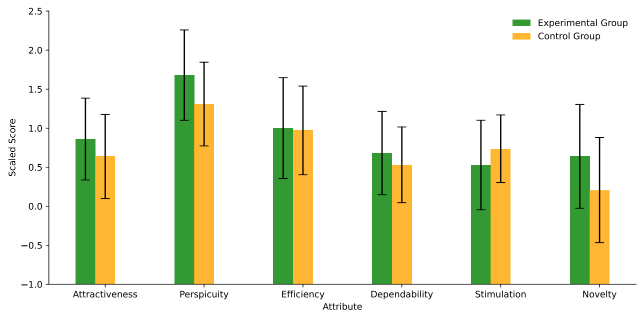
(c) In subplot (c-a), participants in the experimental group exhibited a significantly lower blink rate, as indicated by an asterisks(*). In contrast, (c-b) shows no significant difference was observed in blink duration between the groups.



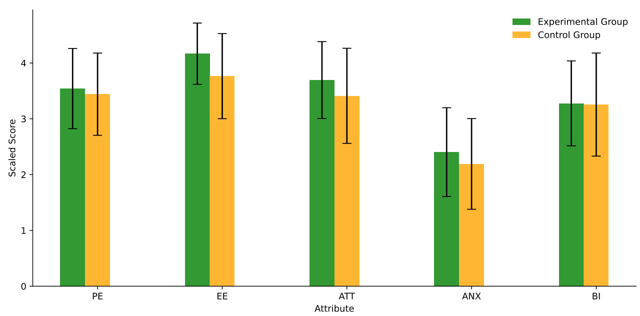
(d) In subplot (d-a), no significant difference found compared averaged saccade rate across the two groups. In subplot (d-b) with asterisks(**), participants in the experimental group exhibited a significantly lower number of average velocity in saccade.

Figure 6: Plots displaying the average pupil dilation (6a), fixation rate and duration (6b), average blink rate and duration (6c), and average saccade rate and velocity (6d) for the experimental and control groups. Accompanying histograms with error bars are also provided for each measure. Attributes and subplots marked with asterisks (* or **) represent significant differences. * denotes 0.01 < p < .05, and ** denotes p < .001.

DesignMinds [40]. Furthermore, the observed increase in gaze fixation duration and faster saccade speed in the experimental group suggests that participants were more engaged in the tasks compared to the control group [16, 34, 76]. Supported by existing literature, longer gaze fixation duration and quicker saccadic movements typically indicate higher levels of focus and cognitive engagement [37, 93]. This may also suggest that our *DesignMinds* captured participants' attention more effectively within the design task context compared to the traditional practice without additional helps in the control group. Similarly, the observed lower blink rate in the experimental group suggests that participants showed greater emotional interest in the generated content, which in turn increased their focus and engagement with the provided design use case [54]. A high level of work engagement has also been shown to lead to more positive and improved work performance [15, 44]. In this way, **participants from the experimental group took the design-specialized advice and engaged in more iterative reflection in the ideation processing**.



(a) Scaled average values for measuring UX (UEQ) which include attributes—Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty—compared between the experimental and control groups.



(b) Scaled average values for measuring technology acceptance and use (UTAUT) across attributes: PE (Performance Expectancy), EE (Effort Expectancy), ATT (Attitude Toward using Technology), ANX (Anxiety), and BI (Behavioral Intention to use the system).

Figure 7: Histograms showcase UX and technology acceptance and use, measured using UEQ and UTAUT questionnaires, respectively. Standard deviations are represented as error bars. No significant main effect was found between the experimental and control groups regarding the introduction of a new type of technology (i.e., *DesignMinds*).

Following the eye-tracking measurements, we conducted a follow-up analysis of the chat logs. We examined how the *DesignMinds*' responses influenced the interactions and how these exchanges correlated with the quality of ideas produced during human divergent thinking. As noted in Section 5.2.2, many participants engaged with *DesignMinds* to seek inspiration and guidance for potential design improvements based on the video context. The video comprehension function from *DesignMinds* augmented the case debriefing process to allow participants to bypass the need to introduce the design case from scratch. Instead, participants could directly propose questions about both general and specific contents from the videos. The conversational interface from *DesignMinds* facilitated ideation through a turn-based flow among design practitioners, domain-specific knowledge base, and video content. This pattern of back-and-forth interaction may have supported reflection during the VBD ideation process. In particular, the interface not only responded to video-derived content but also appeared to encourage some participants to revisit or clarify earlier observations and decisions, as evidenced in parts (b) and (c) in Section 5.2.2. These patterns are broadly consistent with insights from Shaer et al. [74], who found that dialectic interaction with LLMs supported not only idea generation but also the iterative reframing and development of those ideas. While our context and methods differ, we similarly observed (part (d) in Section 5.2.2) that conversational engagement with *DesignMinds* enabled some participants to build incrementally on prior thoughts. In this way, *DesignMinds* may have played a modest role in structuring how ideation unfolded during interaction.

Interestingly, from what we observed in parts (c) and (d) in Section 5.2.2, participants treated *DesignMinds* not merely as a tool but as a collaborator whose responses could validate assumptions based on the knowledge base, introduce novel perspectives, and, in some cases, provoke critical re-evaluation of their ideas. This aligns with Kim et al. [42], who found that overlap-capable chatbots often foster more responsive interactions that enhance users' sense of shared agency in ideation. In our study, some designers utilized *DesignMinds*' contextual understanding from emulating domain expertise to seek confirmation about the use cases or video contents (part (d) in Section 5.2.2). Upon receiving positive feedback, designers became more intrigued and confident which lead to deeper insights during the Divergent Thinking phase. Additionally, some participants incorporated their personal perspectives into the questions and findings they sought to confirm (part (d) in Section 5.2.2). This reflects the nature of design work as emotionally and personally driven [90], and it also aligns with findings by Qin et al. [64], who observed that users engaging with LLMs after forming initial ideas demonstrated higher creative self-efficacy and a stronger sense of ownership over their contributions.

In addition, subsequent correlation tests reveal several strong and positive relationships between the words and ideas generated in chat logs and the quality of ideation in Divergent Thinking tasks. This indicates that ideation outcomes were closely linked to the richness of participants' exchanges with *DesignMinds*. **Consequently, the ideation phase is likely to be enhanced by richer contents from generative answers in *DesignMinds*.** Similarly, prior research has demonstrated that well-structured instructions in design tasks can play a significant role in eliciting higher levels of originality and fostering a broader range of ideation among practitioners [70]. Consistent with this, we observed a positive correlation between the length of participants' questions and the originality of their responses. This suggests that the quality and quantity of the generated answers may be influenced by the level of detail in the query input. This finding emphasizes that it is important to ensure design practitioners to clearly explain their needs in the context of the current circumstances in future studies.

6.3 No Decline in UX or Technology Acceptance and Use with the Introduction of New Technology

When introducing new technology into existing practices, practitioners may struggle with the adaptation process. Technostress, for example, is a phenomenon where individuals are unable to work with new information and communication technologies (ICT) in their work [80]. This difficulty can lead to a decrease in productivity and creativity [10]. Previous literature has shown that discomfort with newly introduced tools often manifests as a decline in UX and in the ratings of technology acceptance and usage [41, 79, 81]. Such a decline can potentially lead to ineffective use of the new technology and mismeasurement of its actual functionality. Given that the *DesignMinds* integrates emerging ICT components, we are particularly interested in understanding whether the prototype affects UX and technology acceptance and use scores compared to the baseline (RQ3). In Section 5.3, the analysis of self-reported scores from two separate questionnaires revealed no significant differences between the experimental and control groups. This suggests that participants in both groups exhibited similar levels of task satisfaction and willingness to accept and use the prototype. **As such, the proposed *DesignMinds* did not negatively impact the normal design ideation experience and did not alter the original use and acceptance of the technology**. Additionally, while we observed lower ratings for certain attributes, such as perceived dependability, stimulation, and novelty within the user experience, these variations do not impact our overall findings of no significant difference in attribute scores. This may be attributable to individual attitudes towards the selected design scenarios, as design is inherently influenced by sentiments and emotions. We anticipate that future studies involving different VBD use cases may yield higher scores, though the pattern of results is expected to remain consistent.

6.4 Limitations and Future Work

While *DesignMinds* shows significant potential for enhancing ideation in VBD, several limitations warrant further investigation. In informal post-experiment discussions, some participants expressed concerns around transparency and trust when using LLMs in creative processes. One of the primary challenges identified is the risk of "hallucinations," a common issue in AI-driven tools where models provide convincing yet incorrect information [39, 49]. This may increase confidence in creative tasks but can also lead to biased or flawed outcomes [61]. To mitigate this risk, we integrated the RAG mechanism [46] into *DesignMinds*. According to prior literature, RAG helps address the issue of generating inaccurate information by enabling the system to retrieve and incorporate task-centric, contextual-relevant, and factual-grounded content [75]. In the future work, we aim to further enhance *DesignMinds*' transparency by integrating more interpretable outputs, such as providing citation links to credible literature in answers [48] which allow designers to trace the rationale behind generated suggestions.

Another limitation is the need to test *DesignMinds* across a broader range of VBD use cases. While *DesignMinds* proved effective in assisting design ideation within the two specific contexts of cooking and construction, real-world applications involve a much wider diversity of design tasks that may demand more flexible tools and an expanded knowledge base. In this study, we predefined the design books for *DesignMinds* 's knowledge repository based on selections made by an independent committee to align with the study's tasks. However, future work could allow designers to personalize the knowledge base by selecting and uploading their own domain-specific resources through a non-programmer-friendly interface. For example, platforms such as AnythingLLM[10] enable users to choose their own LLM models and indexed documents which could potentially offer a more tailored and flexible approach to ideation assistance. Furthermore, our current implementation of ideation assistance offers a fixed level of support to all users. However, design ideation is a highly individualized process, with varying needs for inspiration and suggestions based on the designer's experience [89, 90]. To address this variability, we allowed participants in the study to critically consider their dependability of the assistance according to their own preferences, giving them the freedom to choose which aspects of ideation assistance to utilize and what to record in the Divergent Thinking process. The consistent level of support was maintained to ensure a fair comparison and to isolate *DesignMinds*' impact on ideation. In future iterations, we could consider *DesignMinds* as a product and implement a tunable feature that allows users to adjust the level of "helpfulness" in guiding the design ideation process. We expect this would enable designers to control the amount of information provided according to their needs and makes the tool more responsive to individual preferences.

7 Conclusion

The advancement of generative AI has substantially transformed human work in recent years. In VBD design, there remains an urgent need to reduce the burden of manual video analysis and accelerate professional ideation. Prior research across multiple disciplines has demonstrated efforts to harness the power of generative AI to augment design ideation. In this paper, we present *DesignMinds*, a prototype that elevates ideation assistance for VBD to a higher level. Utilizing advanced techniques from generative AI, our *DesignMinds* can automatically extract information from videos, integrate with professional design guidelines from indexed literature, and provide design- and case-centric recommendations to inspire designers. Our findings demonstrate that *DesignMinds* significantly improves ideation outcomes in terms of flexibility and originality in Divergent Thinking. Through cognitive monitoring via eye-tracking and chat log analysis, we observed increased engagement in design ideation when using *DesignMinds*. Furthermore, assessments of UX and technology acceptance and use indicated that the introduction of this tool did not contribute to increased stress and ensures there will be a smooth integration into the existing VBD workflow in future.

References

[1] Selcuk Acar and Mark A. Runco. 2019. Divergent thinking: New methods, recent research, and extended theory. 13, 2 (2019), 153–158. https://doi.org/10.1037/aca0000231 Place: US Publisher: Educational Publishing Foundation.
[2] Ömer Akin. 1994. Creativity in Design. 7, 3 (1994), 9–21. https://doi.org/10.1111/j.1937-8327.1994.tb00633.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1937-8327.1994.tb00633.x.
[3] Samuel Attuquayefio and Hillar Addo. 2014. Using the UTAUT model to analyze students' ICT adoption. 10, 3 (2014). https://www.learntechlib.org/p/148478/ Publisher: Open Campus, The University of the West Indies, West Indies.
[4] Hasan Ayaz, Patricia A. Shewokis, Scott Bunce, Kurtulus Izzetoglu, Ben Willems, and Banu Onaral. 2012. Optical brain monitoring for operator training and mental workload assessment. 59, 1 (2012), 36–47. https://doi.org/10.1016/j.neuroimage.2011.06.023

[10] https://anythingllm.com/ (last accessed: May 25, 2025)

[5] Fabio Babiloni. 2019. Mental Workload Monitoring: New Perspectives from Neuroscience. In *Human Mental Workload: Models and Applications* (Cham) *(Communications in Computer and Information Science)*, Luca Longo and Maria Chiara Leva (Eds.). Springer International Publishing, 3–19. https://doi.org/10.1007/978-3-030-32423-0_1
[6] John Baer. 2014. *Creativity and Divergent Thinking: A Task-Specific Approach*. Psychology Press. https://doi.org/10.4324/9781315806785
[7] Jackson Beatty. 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin* 91, 2 (1982), 276.
[8] Florian Bordes, Richard Yuanzhe Pang, Anurag Ajay, Alexander C. Li, Adrien Bardes, Suzanne Petryk, Oscar Mañas, Zhiqiu Lin, Anas Mahmoud, Bargav Jayaraman, Mark Ibrahim, Melissa Hall, Yunyang Xiong, Jonathan Lebensold, Candace Ross, Srihari Jayakumar, Chuan Guo, Diane Bouchacourt, Haider Al-Tahan, Karthik Padthe, Vasu Sharma, Hu Xu, Xiaoqing Ellen Tan, Megan Richards, Samuel Lavoie, Pietro Astolfi, Reyhane Askari Hemmat, Jun Chen, Kushal Tirumala, Rim Assouel, Mazda Moayeri, Arjang Talattof, Kamalika Chaudhuri, Zechun Liu, Xilun Chen, Quentin Garrido, Karen Ullrich, Aishwarya Agrawal, Kate Saenko, Asli Celikyilmaz, and Vikas Chandra. 2024-05-27. An Introduction to Vision-Language Modeling. https://doi.org/10.48550/arXiv.2405.17247 arXiv:2405.17247 [cs]
[9] Donald J. Campbell. 1988. Task Complexity: A Review and Analysis. 13, 1 (1988), 40–52. https://doi.org/10.2307/258353 Publisher: Academy of Management.
[10] Shalini Chandra, Anuragini Shirish, and Shirish Srivastava. 2019. Does Technostress Inhibit Employee Innovation? Examining the Linear and Curvilinear Influence of Technostress Creators. 44, 1 (2019). https://doi.org/10.17705/1CAIS.04419
[11] Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, et al. 2024. A survey on evaluation of large language models. *ACM transactions on intelligent systems and technology* 15, 3 (2024), 1–45.
[12] Li-Yuan Chiou, Peng-Kai Hung, Rung-Huei Liang, and Chun-Teng Wang. 2023. Designing with AI: An exploration of co-ideation with image generators. In *Proceedings of the 2023 ACM designing interactive systems conference*. 1941–1954.
[13] Bo T. Christensen and Linden J. Ball. 2016. Dimensions of creative evaluation: Distinct design and reasoning strategies for aesthetic, functional and originality judgments. 45 (2016), 116–136. https://doi.org/10.1016/j.destud.2015.12.005
[14] Vicente Chulvi, Elena Mulet, Amaresh Chakrabarti, Belinda López-Mesa, and Carmen González-Cruz. 2012. Comparison of the degree of creativity in the design outcomes using different design methods. 23, 4 (2012), 241–269. https://doi.org/10.1080/09544828.2011.624501 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/09544828.2011.624501.
[15] Carleton Coffrin, Linda Corrin, Paula de Barba, and Gregor Kennedy. 2014. Visualizing patterns of student engagement and performance in MOOCs. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge* (New York, NY, USA) *(LAK '14)*. Association for Computing Machinery, 83–92. https://doi.org/10.1145/2567574.2567586
[16] John Colombo and D. Wayne Mitchell. 2014. Fixation Location and Fixation Duration as Indices of Cognitive Processing. In *Individual Differences in Early Visual Attention: Fixation Time and Information Processing*. Psychology Press.
[17] Barbara Couture. 1986. Effective ideation in written text: A functional approach to clarity and exigence. *Faculty Publications–Department of English* (1986), 67.
[18] Nigel Cross. 2004. Expertise in design: an overview. 25, 5 (2004), 427–441. https://doi.org/10.1016/j.destud.2004.06.002
[19] Shanna R Daly, Colleen M Seifert, Seda Yilmaz, and Richard Gonzalez. 2016. Comparing ideation techniques for beginning designers. *Journal of Mechanical Design* 138, 10 (2016), 101108.
[20] Rikke Friis Dam and Teo Yu Siang. 2017. What is ideation–and how to prepare for ideation sessions. *Interaction Design Foundation* (2017).
[21] S. Dazkir, Jennifer Mower, Kelly reddy best, and Elaine Pedersen. 2013. An exploration of design students' inspiration process. 47 (2013), 404.
[22] Stefano Delle Monache, Rob Moleman, and Elif Ozcan. 2024. Soundstorm, a collaborative ideation game for sound-driven design. In *Proceedings of the 19th International Audio Mostly Conference: Explorations in Sonic Cultures*. 479–486.
[23] Andrew T Duchowski, Krzysztof Krejtz, Izabela Krejtz, Cezary Biele, Anna Niedzielska, Peter Kiefer, Martin Raubal, and Ioannis Giannopoulos. 2018. The index of pupillary activity: Measuring cognitive load vis-à-vis task difficulty with pupil oscillation. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.
[24] Charles Eastman. 2001. New Directions in Design Cognition: Studies of Representation and Recall. (2001).
[25] Claudia Eckert and Martin Stacey. 2000. Sources of inspiration: a language of design. 21, 5 (2000), 523–538. https://doi.org/10.1016/S0142-694X(00)00022-3
[26] Ana Fucs, Juliana Jansen Ferreira, Vinícius C. V. B. Segura, Beatriz de Paulo, Rogerio Abreu De Paula, and Renato Cerqueira. 2020. Sketch-based Video A Storytelling for UX Validation in AI Design for Applied Research. *CHI Extended Abstracts* (2020). https://doi.org/10.1145/3334480.3375221
[27] Milene Gonçalves, Carlos Cardoso, and Petra Badke-Schaub. [n. d.]. Inspiration choices that matter: the selection of external stimuli during ideation. 2 ([n. d.]), e10. https://doi.org/10.1017/dsj.2016.10
[28] Kristen Grauman, Andrew Westbury, Eugene Byrne, Zachary Chavis, Antonino Furnari, Rohit Girdhar, Jackson Hamburger, Hao Jiang, Miao Liu, Xingyu Liu, Miguel Martin, Tushar Nagarajan, Ilija Radosavovic, Santhosh Kumar Ramakrishnan, Fiona Ryan, Jayant Sharma, Michael Wray, Mengmeng Xu, Eric Zhongcong Xu, Chen Zhao, Siddhant Bansal, Dhruv Batra, Vincent Cartillier, Sean Crane, Tien Do, Morrie Doulaty, Akshay Erapalli, Christoph Feichtenhofer, Adriano Fragomeni, Qichen Fu, Abrham Gebreselasie, Cristina Gonzalez, James Hillis, Xuhua Huang, Yifei Huang, Wenqi Jia, Weslie Khoo, Jachym Kolar, Satwik Kottur, Anurag Kumar, Federico Landini, Chao Li, Yanghao Li, Zhenqiang Li, Karttikeya Mangalam, Raghava Modhugu, Jonathan Munro, Tullie Murrell, Takumi Nishiyasu, Will Price, Paola Ruiz Puentes, Merey Ramazanova, Leda Sari, Kiran Somasundaram, Audrey Southerland, Yusuke Sugano, Ruijie Tao, Minh Vo, Yuchen Wang, Xindi Wu, Takuma Yagi, Ziwei Zhao, Yunyi Zhu, Pablo Arbelaez, David Crandall, Dima Damen, Giovanni Maria Farinella, Christian Fuegen, Bernard Ghanem, Vamsi Krishna Ithapu, C. V. Jawahar, Hanbyul Joo, Kris Kitani, Haizhou Li, Richard Newcombe, Aude Oliva, Hyun Soo Park, James M. Rehg, Yoichi Sato, Jianbo Shi, Mike Zheng Shou, Antonio Torralba, Lorenzo Torresani, Mingfei Yan, and Jitendra Malik. 2021. *Ego4D: Around the World in 3,000 Hours of Egocentric Video*. https://arxiv.org/abs/2110.07058v3
[29] Joshua Guetzkow, Michèle Lamont, and Grégoire Mallard. 2004. What is Originality in the Humanities and the Social Sciences? 69, 2 (2004), 190–212. https://doi.org/10.1177/000312240406900203 Publisher: SAGE Publications Inc.
[30] J. P. Guilford. 1950. Creativity. 5, 9 (1950), 444–454. https://doi.org/10.1037/h0063487 Place: US Publisher: American Psychological Association.
[31] J. P. Guilford. 1967. Creativity: Yesterday, Today and Tomorrow. 1, 1 (1967), 3–14. https://doi.org/10.1002/j.2162-6057.1967.tb00002.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.2162-6057.1967.tb00002.x.
[32] John M. Hinson, Tina L. Jameson, and Paul Whitney. 2003. Impulsive decision making and working memory. 29, 2 (2003), 298–306. https://doi.org/10.1037/0278-7393.29.2.298
[33] Bert Hoeks and Willem J. M. Levelt. 1993-03-01. Pupillary dilation as a measure of attention: a quantitative system analysis. 25, 1 (1993-03-01), 16–26. https://doi.org/10.3758/BF03204445
[34] James E. Hoffman and Baskaran Subramaniam. 1995. The role of visual attention in saccadic eye movements. 57, 6 (1995), 787–795. https://doi.org/10.3758/BF03206794
[35] Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. 2011. *Eye tracking: A comprehensive guide to methods and measures*. oup Oxford.
[36] Shaohan Huang, Li Dong, Wenhui Wang, Yaru Hao, Saksham Singhal, Shuming Ma, Tengchao Lv, Lei Cui, Owais Khan Mohammed, Barun Patra, et al. 2023. Language is not all you need: Aligning perception with language models. *Advances in Neural Information Processing Systems* 36 (2023), 72096–72109.
[37] David E. Irwin. 2004. Fixation Location and Fixation Duration as Indices of Cognitive Processing. In *The Interface of Language, Vision, and Action*. Psychology Press. Num Pages: 29.
[38] David G. Jansson and Steven M. Smith. 1991. Design fixation. 12, 1 (1991), 3–11. https://doi.org/10.1016/0142-694X(91)90003-F
[39] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023-03-03. Survey of Hallucination in Natural Language Generation. 55, 12 (2023-03-03), 248:1–248:38. https://doi.org/10.1145/3571730
[40] Olivia E. Kang, Katherine E. Huffer, and Thalia P. Wheatley. 2014. Pupil Dilation Dynamics Track Attention to High-Level Information. 9, 8 (2014), e102463. https://doi.org/10.1371/journal.pone.0102463 Publisher: Public Library of Science.
[41] Zuheir N. Khlaif, Mageswaran Sanmugam, and Abedulkarim Ayyoub. 2023. Impact of Technostress on Continuance Intentions to Use Mobile Technology. 32, 2 (2023), 151–162. https://doi.org/10.1007/s40299-021-00638-x
[42] JiWoo Kim, Minsuk Chang, and JinYeong Bak. 2025. Beyond Turn-taking: Introducing Text-based Overlap into Human-LLM Interactions. *arXiv preprint arXiv:2501.18103* (2025).
[43] Kyungsun Kim, Jeongyun Heo, and Sanghoon Jeong. 2021. Tool or Partner: The Designer's Perception of an AI-Style Generating Service. In *Artificial Intelligence in HCI: Second International Conference, AI-HCI 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings* (Berlin, Heidelberg). Springer-Verlag, 241–259. https://doi.org/10.1007/978-3-030-77772-2_16
[44] Woocheol Kim, Judith A. Kolb, and Taesung Kim. 2013. The Relationship Between Work Engagement and Performance: A Review of Empirical Literature and a Proposed Research Agenda. 12, 3 (2013), 248–276. https://doi.org/10.1177/1534484312461635 Publisher: SAGE Publications.
[45] Simon Laing and Masood Masoodian. 2016. A study of the influence of visual imagery on graphic design ideation. *Design Studies* 45 (2016), 187–209.
[46] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In *Proceedings of the 34th International Conference on Neural Information Processing Systems* (Vancouver, BC, Canada)

(NIPS'20). Curran Associates Inc., Red Hook, NY, USA, Article 793, 16 pages.
[47] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. 2023. BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models. In *Proceedings of the 40th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 202)*, Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (Eds.). PMLR, 19730–19742. https://proceedings.mlr.press/v202/li23q.html
[48] Weitao Li, Junkai Li, Weizhi Ma, and Yang Liu. 2024. Citation-Enhanced Generation for LLM-based Chatbots. https://doi.org/10.48550/arXiv.2402.16063 arXiv:2402.16063 [cs]
[49] Stephanie Lin, Jacob Hilton, and Owain Evans. 2022. TruthfulQA: Measuring How Models Mimic Human Falsehoods. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Dublin, Ireland), Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (Eds.). Association for Computational Linguistics, 3214–3252. https://doi.org/10.18653/v1/2022.acl-long.229
[50] J. S. Linsey, I. Tseng, K. Fu, J. Cagan, K. L. Wood, and C. Schunn. 2010. A Study of Design Fixation, Its Mitigation and Perception in Engineering Design Faculty. 132, 41003 (2010). https://doi.org/10.1115/1.4001110
[51] Belinda López-Mesa, Elena Mulet, Rosario Vidal, and Graham Thompson. 2011. Effects of additional stimuli on idea-finding in design teams. 22, 1 (2011), 31–54. https://doi.org/10.1080/09544820902911366 Publisher: Taylor & Francis.
[52] Wendy E. Mackay and Anne Laure Fayard. 1999. Video brainstorming and prototyping: techniques for participatory design. In *CHI '99 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA) *(CHI EA '99)*. Association for Computing Machinery, 118–119. https://doi.org/10.1145/632716.632790
[53] Wendy E. Mackay, Anne V. Ratzer, and Paul Janecek. 2000. Video artifacts for design: bridging the Gap between abstraction and detail. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques* (New York City New York USA). ACM, 72–82. https://doi.org/10.1145/347642.347666
[54] Antonio Maffei and Alessandro Angrilli. 2019. Spontaneous blink rate as an index of attention and emotion during film clips viewing. 204 (2019), 256–263. https://doi.org/10.1016/j.physbeh.2019.02.037
[55] Liane Makatura, Michael Foshey, Bohan Wang, Felix HähnLein, Pingchuan Ma, Bolei Deng, Megan Tjandrasuwita, Andrew Spielberg, Crystal Elaine Owens, Peter Yichen Chen, Allan Zhao, Amy Zhu, Wil J. Norton, Edward Gu, Joshua Jacob, Yifei Li, Adriana Schulz, and Wojciech Matusik. 2023. How Can Large Language Models Help Humans in Design and Manufacturing? https://doi.org/10.48550/arXiv.2307.14377 arXiv:2307.14377 [cs]
[56] Philip B Meggs and Alston W Purvis. 2016. *Meggs' history of graphic design.* John Wiley & Sons.
[57] Humza Naveed, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Naveed Akhtar, Nick Barnes, and Ajmal Mian. 2023. A comprehensive overview of large language models. *arXiv preprint arXiv:2307.06435* (2023).
[58] OpenAI. 2023. GPT-4 Technical Report. arXiv:2303.08774 [cs.CL]
[59] OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks, Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen, Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung, Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch, Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte, Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott Gray, Ryan Greene, Joshua Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade Hickey, Peter Hoeschele, Brandon Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shawn Jain, Joanne Jang, Angela Jiang, Roger Jiang, Haozhun Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo Jun, Tomer Kaftan, Łukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar, Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Jan Hendrik Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Łukasz Kondraciuk, Andrew Kondrich, Aris Konstantinidis, Kyle Kosic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie Lin, Mateusz Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini, Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie Mayer, Andrew Mayne, Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély, Ashvin Nair, Reiichiro Nakano, Rajeev Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo Noh, Long Ouyang, Cullen O'Keefe, Jakub Pachocki, Alex Paino, Joe Palermo, Ashley Pantuliano, Giambattista Parascandolo, Joel Parish, Emy Parparita, Alex Passos, Mikhail Pavlov, Andrew Peng, Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira Pinto, Michael, Pokorny, Michelle Pokrass, Vitchyr H. Pong, Tolly Powell, Alethea Power, Boris Power, Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond, Francis Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario Saltarelli, Ted Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John Schulman, Daniel Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav Shyam, Szymon Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama, Ian Sohl, Benjamin Sokolowsky, Yang Song, Natalie Staudacher, Felipe Petroski Such, Natalie Summers, Ilya Sutskever, Jie Tang, Nikolas Tezak, Madeleine B. Thompson, Phil Tillet, Amin Tootoonchian, Elizabeth Tseng, Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea Vallone, Arun Vijayvergiya, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben Wang, Jonathan Ward, Jason Wei, C. J. Weinmann, Akila Welihinda, Peter Welinder, Jiayi Weng, Lilian Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong, Lauren Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu, Qiming Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao, Tianhao Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. 2024. GPT-4 Technical Report. https://doi.org/10.48550/arXiv.2303.08774 arXiv:2303.08774 [cs]
[60] Laurence Paquette and Thomas Kida. 1988. The effect of decision strategy and task complexity on decision performance. 41, 1 (1988), 128–142. https://doi.org/10.1016/0749-5978(88)90051-9
[61] Raja Parasuraman and Dietrich H. Manzey. 2010. Complacency and Bias in Human Use of Automation: An Attentional Integration. 52, 3 (2010), 381–410. https://doi.org/10.1177/0018720810376055 Publisher: SAGE Publications Inc.
[62] Zhiliang Peng, Wenhui Wang, Li Dong, Yaru Hao, Shaohan Huang, Shuming Ma, and Furu Wei. 2023. Kosmos-2: Grounding multimodal large language models to the world. *arXiv preprint arXiv:2306.14824* (2023).
[63] Cyril Picard, Kristen M. Edwards, Anna C. Doris, Brandon Man, Giorgio Giannone, Md Ferdous Alam, and Faez Ahmed. 2023. From Concept to Manufacturing: Evaluating Vision-Language Models for Engineering Design. arXiv:2311.12668 [cs] http://arxiv.org/abs/2311.12668
[64] Peinuan Qin, Chi-Lan Yang, Jingshu Li, Jing Wen, and Yi-Chieh Lee. 2025. Timing Matters: How Using LLMs at Different Timings Influences Writers' Perceptions and Ideation Outcomes in AI-Assisted Ideation. *arXiv preprint arXiv:2502.06197* (2025).
[65] Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin* 124, 3 (1998), 372.
[66] Jeba Rezwana and Mary Lou Maher. 2023. Designing Creative AI Partners with COFI: A Framework for Modeling Interaction in Human-AI Co-Creative Systems. 30, 5 (2023), 1–28. https://doi.org/10.1145/3519026
[67] Mark A. Runco. 2010. Chapter 12: Divergent Thinking, Creativity and Ideation. In *The Cambridge Handbook of Creativity*. Cambridge University Press.
[68] Mark A. Runco and Selcuk Acar. 2024. Divergent Thinking as an Indicator of Creative Potential. 24, 1 (2024), 66–75. https://doi.org/10.1080/10400419.2012.652929 Publisher: Routledge _eprint: https://doi.org/10.1080/10400419.2012.652929.
[69] Mark A. Runco and Shawn M. Okuda. 1988. Problem discovery, divergent thinking, and the creative process. 17, 3 (1988), 211–220. https://doi.org/10.1007/BF01538162
[70] M. A. Runco and S. M. Okuda. 1991. The instructional enhancement of the flexibility and originality scores of divergent thinking tests. 5, 5 (1991), 435–441. https://doi.org/10.1002/acp.2350050505 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/acp.2350050505.
[71] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2014. Applying the User Experience Questionnaire (UEQ) in Different Evaluation Scenarios. 383–392. https://doi.org/10.1007/978-3-319-07668-3_37
[72] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Construction of a Benchmark for the User Experience Questionnaire (UEQ). 4 (2017), 40–44. https://doi.org/10.9781/ijimai.2017.445
[73] Rossitza Setchi and Carole Bouchard. 2010. In Search of Design Inspiration: A Semantic-Based Approach. 10, 31006 (2010). https://doi.org/10.1115/1.3482061
[74] Orit Shaer, Angelora Cooper, Osnat Mokryn, Andrew L Kun, and Hagit Ben Shoshan. 2024. AI-Augmented Brainwriting: Investigating the use of LLMs in group ideation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–17.
[75] Kurt Shuster, Spencer Poff, Moya Chen, Douwe Kiela, and Jason Weston. 2021. Retrieval Augmentation Reduces Hallucination in Conversation. https://doi.org/10.48550/arXiv.2104.07567 arXiv:2104.07567 [cs]
[76] Miriam Spering. 2022. Eye Movements as a Window into Decision-Making. 8 (2022), 427–448. Issue Volume 8, 2022. https://doi.org/10.1146/annurev-vision-

100720-125029 Publisher: Annual Reviews.
[77] John A Stern, Larry C Walrath, and Robert Goldstein. 1984. The endogenous eyeblink. *Psychophysiology* 21, 1 (1984), 22–33.
[78] Catherine Stones and Tom Cassidy. 2010. Seeing and discovering: how do student designers reinterpret sketches and digital marks during graphic design ideation? *Design studies* 31, 5 (2010), 439–460.
[79] Monideepa Tarafdar, Ellen Bolman. Pullins, and T. S. Ragu-Nathan. 2015. Technostress: negative effect on performance and possible mitigations. 25, 2 (2015), 103–132. https://doi.org/10.1111/isj.12042 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/isj.12042.
[80] Monideepa Tarafdar, Qiang Tu, Bhanu S. Ragu-Nathan, and T. S. Ragu-Nathan. 2007. The Impact of Technostress on Role Stress and Productivity. 24, 1 (2007), 301–328. https://doi.org/10.2753/MIS0742-1222240109 Publisher: Routledge _eprint: https://doi.org/10.2753/MIS0742-1222240109.
[81] Monideepa Tarafdar, Qiang Tu, and T. S. Ragu-Nathan. 2010. Impact of Technostress on End-User Satisfaction and Performance. 27, 3 (2010), 303–334. https://doi.org/10.2753/MIS0742-1222270311 Publisher: Routledge _eprint: https://doi.org/10.2753/MIS0742-1222270311.
[82] Deborah Tatar. 1989. Using video-based observation to shape the design of a new technology. *ACM SIGCHI Bulletin* 21, 2 (1989), 108–111. https://doi.org/10.1145/70609.70628
[83] Viswanath Venkatesh, Michael G. Morris, Gordon B. Davis, and Fred D. Davis. 2003. User Acceptance of Information Technology: Toward a Unified View. https://papers.ssrn.com/abstract=3375136
[84] Laurie Vertelney. 1989. Using video to prototype user interfaces. *ACM SIGCHI Bulletin* 21, 2 (1989), 57–61. https://doi.org/10.1145/70609.70615
[85] Vimal K. Viswanathan and Julie S. Linsey. 2013. Design Fixation and Its Mitigation: A Study on the Role of Expertise. 135, 51008 (2013). https://doi.org/10.1115/1.4024123
[86] Michael D Williams, Nripendra P Rana, and Yogesh K Dwivedi. 2015. The unified theory of acceptance and use of technology (UTAUT): a literature review. 28, 3 (2015), 443–488. https://doi.org/10.1108/JEIM-09-2014-0088 Publisher: Emerald Group Publishing Limited.
[87] Shengyang Xu, Yao Wei, Pai Zheng, Jia Zhang, and Chunyang Yu. 2024. LLM enabled generative collaborative design in a mixed reality environment. 74 (2024), 703–715. https://doi.org/10.1016/j.jmsy.2024.04.030
[88] Xiaotong (Tone) Xu, Jiayu Yin, Catherine Gu, Jenny Mar, Sydney Zhang, Jane L. E, and Steven P. Dow. 2024. Jamplate: Exploring LLM-Enhanced Templates for Idea Reflection. In *Proceedings of the 29th International Conference on Intelligent User Interfaces* (New York, NY, USA) *(IUI '24)*. Association for Computing Machinery, 907–921. https://doi.org/10.1145/3640543.3645196
[89] Salu Ylirisku and Jacob Buur. 2007. Making sense and editing videos. In *Designing with video: Focusing the user-centred design process*. Springer London, 86—-135. https://doi.org/10.1007/978-1-84628-961-3_2
[90] Salu Ylirisku and Jacob Buur. 2007. Studying what people do. In *Designing with video: Focusing the user-centred design process*. Springer London, 36–85. https://doi.org/10.1007/978-1-84628-961-3_2
[91] Robert J. Youmans and Thomaz Arciszewski. 2014. Design fixation: Classifications and modern methods of prevention. 28, 2 (2014), 129–137. https://doi.org/10.1017/S0890060414000043
[92] Carmen Zahn, Roy Pea, Friedrich W. Hesse, and Joe Rosen. 2010. Comparing Simple and Advanced Video Tools as Supports for Complex Collaborative Design Processes. 19, 3 (2010), 403–440. https://doi.org/10.1080/10508401003708399 Publisher: Routledge _eprint: https://doi.org/10.1080/10508401003708399.
[93] Min Zhao, Timothy M. Gersch, Brian S. Schnitzer, Barbara A. Dosher, and Eileen Kowler. 2012. Eye movements and attention: The role of pre-saccadic shifts of attention in perception, memory and the control of saccades. 74 (2012), 40–60. https://doi.org/10.1016/j.visres.2012.06.017
[94] Chuyi Zhou, Xiyuan Zhang, and Chunyang Yu. 2024. How does AI promote design iteration? The optimal time to integrate AI into the design process. 0, 0 (2024), 1–28. https://doi.org/10.1080/09544828.2023.2290915 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/09544828.2023.2290915.
[95] Gengze Zhou, Yicong Hong, and Qi Wu. 2023. NavGPT: Explicit Reasoning in Vision-and-Language Navigation with Large Language Models. https://doi.org/10.48550/arXiv.2305.16986 arXiv:2305.16986 [cs]
[96] J. Zimmerman. 2005. Video sketches: exploring pervasive computing interaction designs. *IEEE Pervasive Computing* 4, 4 (2005), 91–94. https://doi.org/10.1109/MPRV.2005.91

You will be shown two videos. Your task is to analyze the videos and pinpoint processes or methods that could be enhanced. Focus on the activities and consider alternative tools, interactions, or contextual improvements. Generate and write out as many ideas as possible. You are encouraged to think out loud.

(Please use the provided chatbot to assist you. This tool offers insights and suggests improvements based on the video content. Type your questions or thoughts into the chatbot and use its responses to enhance your ideation. For example, ask, "How can the process shown in the video be improved?")

You will have 15 minutes to engage with each video. Please use your time effectively and document as many ideas as possible. Please note that videos do not have sound. You will be notified after 12 minutes of the time.

When you are ready to proceed press the "Start" button and the arrow "→" on the bottom right side of the screen.

Supplementary Text 1: Instructional text displayed in the Note-taking Space in Fig. 3. Text within parentheses (the second paragraph) was shown only to participants in the experimental group with access to *DesignMinds*.

- *Flexibility: Each comprehensive idea which portraying the purpose and functionality in sufficient detail to be understandable gives a +1 point.*
- *Flexibility: Give a +1 point for each new domain/subdomain is spotted based on the ideation context across all participants.*
- *Originality: A grade based on the statistical infrequency of ideas measured on a 7-point Likert scale.*

Supplementary Text 2: Predetermined criteria based on Guilford's study [30] for evaluating fluency, flexibility, and originality in divergent thinking texts by independent raters.