DesignAID: Using Generative AI and Semantic Diversity for Design Inspiration

Alice Cai
acai@college.harvard.edu
Harvard University
Cambridge, Massachusetts, USA

Steven R. Rick
srick@mit.edu
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA

Jennifer Heyman
heymanj@mit.edu
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA

Yanxia Zhang
yanxia.zhang@tri.global
Toyota Research Institute
Los Altos, California, USA

Alexandre Filipowicz
alex.filipowicz@tri.global
Toyota Research Institute
Los Altos, California, USA

Matthew K. Hong
matt.hong@tri.global
Toyota Research Institute
Los Altos, California, USA

Matthew Klenk
matt.klenk@tri.global
Toyota Research Institute
Los Altos, California, USA

Thomas Malone
malone@mit.edu
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA

Figure 1: Two mood boards created by participants using our generative AI system for creative ideation (left) and a Pinterest search system (right), in response to the prompt "chair for children".

ABSTRACT

Designers often struggle to sufficiently explore large design spaces, which can lead to design fixation and suboptimal outcomes. Here we introduce *DesignAID*, a generative AI tool that supports broader design space exploration by first using large language models to produce a range of diverse ideas expressed in words, and then using image generation software to create images from these words. This innovative combination of AI-based capabilities allows human-computer pairs to rapidly create a diverse set of visual concepts without time-consuming drawing. In a study with 87 crowd-sourced designers, we found that designers rated the automatic generation of images from words as significantly more inspirational, enjoyable, and useful than a conventional baseline condition of image search using Pinterest. Surprisingly, however, we found that automatically generating highly diverse ideas had less value. For image generation, the high diversity condition was somewhat better in inspiration but no better in the other dimensions, and for image search it was significantly worse in all dimensions.

CI '23, November 06–09, 2023, Delft, Netherlands

ACM ISBN 979-8-4007-0113-9/23/11.
https://doi.org/10.1145/3582269.3615596

CCS CONCEPTS

• **Computing methodologies → Artificial intelligence**; • **Applied computing**; • **Human-centered computing** → *Human computer interaction (HCI)*; HCI design and evaluation methods; Interaction paradigms;

KEYWORDS

creativity support, human-computer collaboration, human AI collaboration, AI assistance, generative AI, machine learning

ACM Reference Format:
Alice Cai, Steven R. Rick, Jennifer Heyman, Yanxia Zhang, Alexandre Filipowicz, Matthew K. Hong, Matthew Klenk, and Thomas Malone. 2023. *DesignAID*: Using Generative AI and Semantic Diversity for Design Inspiration. In *Collective Intelligence Conference (CI '23), November 06–09, 2023, Delft, Netherlands.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3582269.3615596

1 INTRODUCTION

An important kind of collective intelligence arises in groups consisting of one person and one computer. One important kind of task such groups perform is designing visual and physical objects. In this paper, we explore how generative AI tools can help human-AI groups do this kind of design.

The visual design process is a complex task, requiring imaginative exploration of a broad landscape of possible designs. But designers often face difficulties in doing this because of *design fixation*, the tendency to become focused on a limited set of solutions to a problem [21]. This, combined with our natural tendency to ideate around what is familiar to us [52], can lead to prematurely converging on suboptimal solutions instead of broadly exploring a larger space of possibilities. Another challenge in visual design is that it usually takes experts a long time, sometimes hours or days, to sketch out initial sets of ideas. This manual process is lengthy and costly, which hinders rapid experimentation with design ideas.

Recent advances in generative AI such as GPT-3 [5] and DALL-E [38] suggest possible remedies for these problems. Generative AI systems can now rapidly create unique visual and textual content at a level of quality previously only achievable by humans [5, 40, 47]. For instance, these systems have been used to generate creative outputs in a variety of domains, such as fashion, architecture, writing, product design, and interior design [1, 3, 23, 34]. The systems also have the potential to explicitly expand the diversity of ideas designers consider by selecting ideas that are widely distributed in what is called the *semantic embedding space* of ideas [7, 32]. For instance, early work in crowd-based ideation suggests that people generate more diverse ideas after being exposed to semantically diverse sets of idea descriptions [41]. The combination of these text and image models, when paired with human users, enables an innovative form of collective intelligence.

To demonstrate these possibilities, we introduce *DesignAID*, a generative AI tool that assists designers during the early stages of the creative design process: gathering inspiration and developing ideas. We support broader exploration of ideas by automating the creation of a diverse range of verbal phrases from the semantic embedding space and visual images corresponding to these phrases.

More specifically, our system has two modes. In the low-diversity mode (Direct Input), users enter phrases that directly generate images. In the high-diversity mode (New Ideas), the phrases from users generate an additional set of related, but semantically diverse, phrases. Then, each of these diverse phrases is used to generate an image that corresponds to that phrase.

After 87 designers used the system, we found that they felt it made their work more inspiring, enjoyable, and useful relative to a widely used image search tool (Pinterest).

In summary, we present the following contributions:

- *DesignAID*, a software system for creating and selecting design ideas and visual representations thereof.
- Methods for creating diverse design ideas based on maximizing distances between semantic embeddings of AI-generated ideas.
- A study with designers to measure their subjective experience and behavior using image generation and image search in high and low diversity modes during early-stage design ideation.

2 RELATED WORK

We draw on prior research in the psychology of inspiration and the creative process, generative AI, and collective intelligence for creative tasks.

2.1 The Creative Process and Design

The creative process has been studied extensively in psychology, art, design, and education throughout history. Despite its complexity and seeming intangibility, researchers have proposed various frameworks, stages, and cognitive models to gain an understanding of how ideas emerge, evolve, and manifest as creative products. According to Gabora, creativity is the process of forming new combinations or reorganizing existing ones, resulting in originality, novelty, and utility [16]. From Wallas's four stages of creativity (Preparation, Incubation, Illumination, and Verification) [48], to Csikszentmihalyi's and others' adaptations of those stages [11], to Koestler's theory of creative insight arising from the bisociation of cognitive spaces [26], there are numerous models for analyzing the creative process, yet no consensus. Stage-based models often receive criticism due to both their inability to capture co-occurring and recursive subprocesses and the difficulty of operationalizing them [29]. More universal is the agreement that inspiration in particular is an integral part of the creative process that can be correlated with more effective solutions [18].

The creative design process can be modeled by the general creative process, with the added consideration of constraint-based iteration to transform ideas into tangible solutions. Models of the design process usually follow an iterative approach, like the Double Diamond Design Process [10], which proposes divergent 'Discover' and 'Develop' stages, each followed by convergent 'Define' and 'Deliver' stages respectively in the iterative cycle of ideating and refining problems and solutions, as diagrammed in Figure 2.

Early-stage design processes often involve exploratory techniques such as brainstorming and idea generation, aiming to identify requirements and constraints associated with the task at hand. During such early-stage processes, it is important to broadly explore

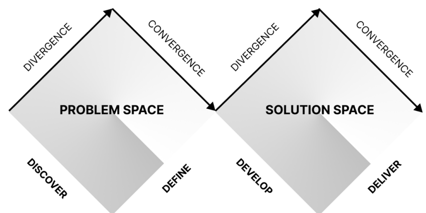

Figure 2: The double diamond design process, comprised of two cycles of divergence and convergence, first in the problem space and then in the solution space.

the design space and generate ideas which may be unconventional yet potentially viable solutions. Diversity in expertise of team members has been shown to produce more novel ideas [20]. However, ideation activities can be long and laborious due to the tedium associated with producing visual representations of these ideas. This, along with mental tendencies such as functional fixedness [17], often leads to design fixation [22].

To address the challenge of design fixation, studies have commonly addressed prevention methods such as concept mapping, remote association, exposure to other new and unrelated content [27, 43], and crowd-based diversity ratings [41]. These approaches, however, still largely rely on human effort and knowledge resources.

Generative AI systems, on the other hand, are prime candidates for assisting human designers in the early stages of the creative process because these systems draw from huge data sets and can complement the human designers' abilities as part of a collectively intelligent system. For example, these systems can rapidly generate content, allowing designers to quickly iterate through ideas and shift their efforts from creating visualizations of all ideas to selecting those which they deem worthy of further exploration and refinement.

However, most previous work on these AI tools focused on acceleration of the content generation process, with few studies examining how AI can be used to assist designers in the early stages of the design process to facilitate divergent thinking and complement the human creative process.

2.2 Generative AI

Recent advancements in generative AI models, such as the impressive gains achieved by the transformer architecture [47], has enabled the development of powerful systems capable of generating image, audio, and text content at a level of complexity previously only achievable by humans. These models are trained on large data sets of images and text and can generate unique content based on learned and reproduced patterns.

Within natural language generation, the release of the 175-billion-parameter language model, GPT-3, in 2020 fueled tremendous research and development efforts within industry, academia, and tech communities across the globe due to its versatility in performing a wide range of tasks [5]. Advances in language models have also led to extraordinary growth in image generation, with systems such as DALL-E 2 [38]. Improving on previous advancements in traditional GAN architectures [51], the development of text-guided diffusion models has allowed for better precision and control in the generation of images from text [15]. Diffusion models work by gradually adding Gaussian noise to training data, then learning to recover that training data from novel noise. Guided by image-text embeddings, such as OpenAI's Contrastive Language-Image Pre-Training (CLIP) network, diffusion models are able to generate high-quality images purely from text instructions [37]. More recently, researchers have found ways to use cross-attention control to more precisely generate images from word level input, paving the way for prompt-to-prompt edits in an unlimited number of variations [19].

In addition to generative AI, semantic embeddings have been used extensively in a variety of AI applications including recommendation systems [44], sentiment analysis [45], and other natural language processing applications. Semantic embeddings are vector-based representations of words and phrases which capture latent contextual meaning as well as shared statistical regularities between words. One example of a popular semantic embedding system is word2vec, which uses neural networks to learn word associations from large corpora of text and represents words as vectors within a high dimensional space [33]. Within the design process, these semantic embeddings can provide an efficient mechanism for generating unique AI-based variations on user input ideas based on their relative positions in the semantic space.

2.3 Collectively Intelligent Human-AI Teams for Creative Tasks

A number of studies have investigated how AI systems can augment human creativity in human-AI teams. For example, Biermann et al. [4] investigated the dynamics between humans and AI writing companions, in which storywriters desired AI companions to respect their personal values and writing strategies. Singh et al. further uncover the human role of performing integrative leaps in incorporating AI-generated text into their developing stories [42].

For visual tasks, FashionQ [23] is an AI creativity support tool for fashion designers that facilitates both convergent and divergent thinking. It does this by providing visualizations that help designers recognize their styles quickly and analytically in a quantitative way. It also helps them study fashion trends across the seasons and combine styles to broaden the extent of ideation. Another work, GANSpiration provides user interface designers with generated images of interfaces for inspiration [34]. It leverages StyleGAN to perform style transfer on a given image and generate new design examples that merge the original input with a random set of existing designs in the training set. Simliarly, We-toon [24] also uses GAN-based image synthesis and manipulation to allow writers to revise any sketch created by artists.

When designing human-AI creative teams, it is important to consider psychological aspects such as control, ownership and trust. As human designers collaborate with AI systems to produce their work, they need to be engaged enough to feel ownership over the output while trusting in the abilities of both themselves and the machine in order to achieve optimal results. One approach toward this balance is to allow users to choose varying degrees of control over the support tool, defining the work they would like to do and the work they would like to offload to the system [14].

AI systems can support the creative process for human users in a number of different ways, including basic support (e.g., smart suggestion and templating systems), analytical analysis (statistical inference and exploratory data analysis), and generative synthesis (automated generation of new solutions based on input models or training datasets). Our work uses generative synthesis methods to identify promising directions and opportunities that designers may not have considered before, increasing the scope and speed for exploration of ideas.

While previous work has explored the dynamics of convergent and divergent modes of thinking within creativity [35], we believe that *DesignAID* is the first to explicitly explore convergent and divergent ideation with generative AI assistance.

3 DESIGNAID

3.1 Design Objectives

Given the complexity of visual design, we focus on the subtask of mood board generation as a proxy for exploring ideation in general. Mood boards are selections of visual and text artifacts curated to evoke particular concepts or aesthetics. Research suggests that mood board design follows an iterative, cyclic process [25]. This process begins with the curation of an image collection, followed by a pruning process that works to refine the board until the desired aesthetic is achieved. Drawing from this research literature, DesignAID has the following objectives:

(1) enabling rapid visualization and iteration of ideas
(2) expanding the operational design space to avoid design fixation
(3) facilitating both creative exploration and refinement through divergent (New Idea) and convergent (Direct Input) modes of operation

3.2 System Design

DesignAID facilitates a user's iterative ideation process by supporting rapid visualization of diverse ideas. To accomplish this, we use a large language model and semantic-distance calculations to generate diverse new ideas and then a text-to-image model to generate associated visual representations. *DesignAID* provides two modes of operation, as shown in Figure 3.

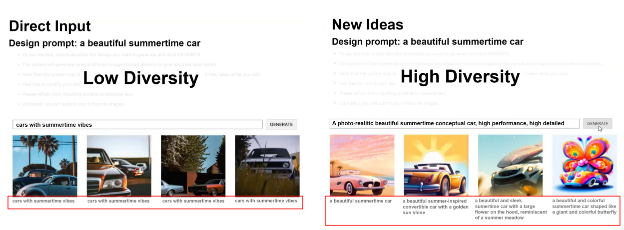

Figure 3: Direct Input and New Idea modes of operation in the *DesignAID* system in action. The user-inputted prompt is used to directly generate images in Direct Input (low diversity) mode (left). The user-inputted prompt is used to generate new ideas which then generate images in the New Idea (high diversity) mode (right).

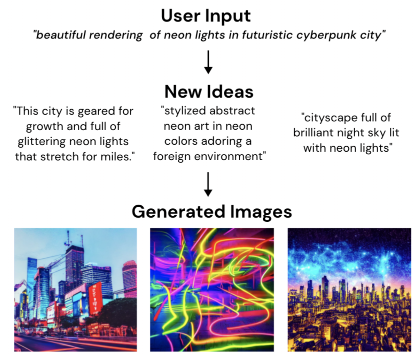

Figure 4: An example set of new ideas generated by *DesignAID* in the New Idea Mode, where the user inputs a prompt and receives a set of related but diverse ideas and images.

Direct Input Mode (low diversity) takes the user's text input and directly uses it to drive the creation of a series of images through a text-to-image generation pipeline. This mode facilitates rapid iteration for refining ideas, as users can change their text input and continue to generate images to explore small or large variations within the design space.

New Idea Mode (high diversity) uses much of the same architecture from our Direct Input mode, but before getting to image generation it first uses a generative language model to generate a large set of new ideas related to the text input entered by the user. These new ideas are evaluated against one another, and the set of the most semantically different ideas are selected and then used to generate images (see Figure 4). This mode aids in the exploration of the design landscape as the new ideas and their associated images are intended to spark the creativity needed to break out of any design fixation.

3.3 Technical Implementation

To address our first objective of enabling rapid ideation, DesignAID is built using a variant of Stable Diffusion, a latent text-to-image diffusion model [40]. This technology was trained on a large corpus of labeled data in order to learn how to deconstruct an image down to random noise and then reconstruct that input image again from the deconstructed random noise. This grants the ability for the model to take an input (random noise image), text target, and some parameters such as number of time steps and then plan out a reconstruction and generate largely unique and novel imagery that represents the target text. We configured a FastAPI [1] endpoint in order to grant scalable access to this text-to-image model. We attached a load-balancer to the API in order to minimize the amount of time each user waited for image generation by distributing work across several AWS EC2 p3.2xlarge instances. The load-balancer ensured active users were allocated to the least busy instance when they requested images be generated through the application. In order to further optimize for performance and output quality we experimented with parameters and found that an NVIDIA Tesla V100 could output 512x512 images in 25 timesteps in approximately 6 seconds with acceptable accuracy and quality. While a greater number

[1]https://fastapi.tiangolo.com/

of timesteps leads to images that better match the input text, it also leads to considerably longer generation times. We elected to restrict this parameter in this study in order to prioritize interactivity and system responsiveness.

We also implement the use of image search to create images within our application in order to evaluate a baseline condition of image search vs. image generation. We search Pinterest for imagery as this is a commonly used resource in the design industry and seemed the closest option to the standard practice for gathering inspiration. This functions as an additional endpoint in our API that uses a GET request that includes user text input as a URL parameter and receives back the top image results from the Pinterest search algorithm.

To address our second objective of avoiding design fixation, *DesignAID* generates a semantically diverse sets of new ideas through an integration with OpenAI's GPT-3 text generation system [5]. GPT, or Generative Pre-trained Transformer, is a large language model that attempts to predict what text should come next given a textual prompt. In our case, we give the system (a) the user's input, (b) instructions to generate new ideas, and (c) a few examples of the kinds of new ideas to generate (see sample prompt in Appendix B). In response to each prompt, GPT-3 generates approximately seven related but semantically different ideas.

We explored prompt engineering and few-shot training in order to guide the output to align with our goals. In particular, we found that loosely repeating structures with high usage of synonymous adjectives and semantically similar terminology encouraged the output to be more wide-ranging and diverse. From this set of diverse generated ideas, the three that are most different from one another are selected using Google's Universal Sentence Encoder semantic embedding computations [8]. In order to select the most diverse set, the system first measures the similarity between each pair of generated ideas by computing the Euclidean distance between available vectors. These semantic distances are then used to select the subset of most diverse ideas based on an incremental furthest search algorithm.

For instance, if a designer inputs "nature-inspired chair", the system generates the following ideas:

- "sleek and modern nature-inspired chair made of bamboo"
- "elegant and abstract nature-inspired lounge chair that looks like a tree"
- "fancy and creative nature-inspired recliner made of stones and pebbles"
- "chic nature-inspired armchair made of colorful leaves"
- "luxurious nature-inspired rocking chair made of wood and twigs"
- "delicate and graceful nature-inspired sofa made of feathers refined nature-inspired seat made of sticks and branches in a wicker weaving pattern"

From these ideas, the most diverse set selected are:

- "sleek and modern nature-inspired chair made of bamboo"
- "fancy and creative nature-inspired recliner made of stones and pebbles"
- "chic nature-inspired armchair made of colorful leaves"

Finally, as in the Direct Input mode, the New Idea mode generates or searches for visual representations of these ideas. This full pipeline is diagrammed in Figure 5.

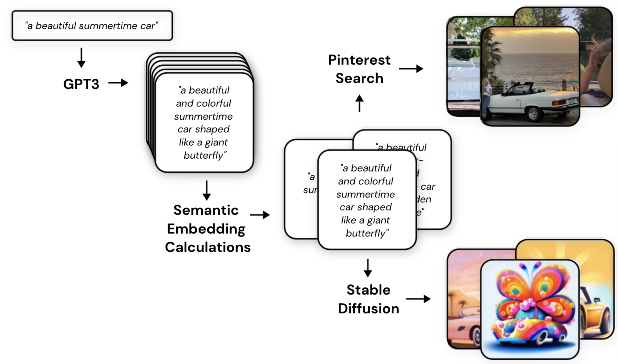

Figure 5: The pipeline from user input to image output in the divergent (New Idea) mode of the tool: (a) the user inputs a phrase (e.g., "a beautiful summertime car"), (b) GPT-3 generates new (longer) phrases based on the user's input, (c) semantic embedding calculations are used to select the most diverse set of these phrases, and (d) Stable Diffusion model generates images from the selected phrases.

3.4 User Interface

To help users take advantage of *DesignAID*, we built a simple and task-focused web application, similar in concept to other preexisting generative image systems like DALL-E 2. *DesignAID* predominantly consists of a text input box, followed by a row of rendered images and their associated prompts, as seen in Figure 6. As each round of idea and image generation concludes, the outputs are added to a collection at the bottom of the application. This collection serves as a space for users to reflect on previously generated ideas and visual outputs as well as a place where they can select the images to include in their final mood board.

In order to allow us to conduct experiments with participants online, we embedded the *DesignAID* application within Empirica, a larger experiment administration framework that allows for condition assignment, controlling experiment timing, and user activity logging [2]. While the Empirica system imposes some constraints on the exact implementation, it generally supports React-based components within a modern web-development framework. This means that *DesignAID* can be easily transitioned into a standalone application at a later time.

4 FORMATIVE EVALUATION

4.1 Methods

To gain initial insights about the usability and effectiveness of the system, we ran a series of pilot studies in which participants were asked to create two mood boards in response to two different prompts. Specifically, participants created one mood board in the New Idea mode and one mood board in the Direct Input mode. After creating each mood board, participants completed a survey that asked open-ended questions about their experience using the *DesignAID* tool for that task.

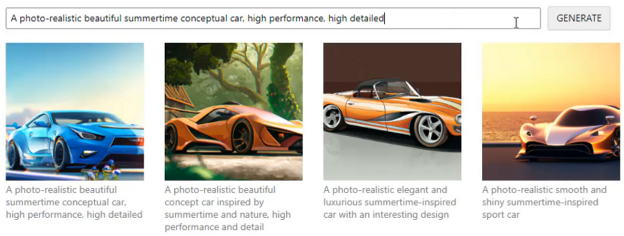

Figure 6: The image generation part of the user interface of *DesignAID*, in the New Idea mode. A text input field allows users to enter a phrase that then generates ideas and visual images of those ideas. Each generated image includes its associated prompt (all images have unique prompts in the New Idea mode whereas all images use the same input prompt for Direct Input mode).

After both mood boards and task surveys were complete, participants completed a final survey where they reflected on their experience across the two treatment conditions.

4.2 Participants

A total of 33 participants were recruited across three pilot studies from Upwork and university student populations. The pilot studies consisted of 5, 8, and 20 participants, respectively.

Of our 33 participants, 28 were successfully able to complete the tasks as described (generate images until satisfied with collection, then curate nine into a single mood board, across two treatment conditions). We attribute the five failures to a combination of user errors and system delays.

4.3 Results

4.3.1 Inspiration. A number of pilot participants considered the New Idea mode to be more inspirational. For instance, Participant 14 (P14) found the AI-generated ideas to be a "key element" during the task. Another participant said that the new ideas "helped me to make associations with other images and search them more efficiently" (P21).

Additionally, the New Idea mode provided images and concepts that the participants may not have thought of on their own, and thus opened up a new avenue of creative thinking for them. One participant stated that they were "quite pleasantly surprised" by the images that were generated in response to their prompt (P19), while another participant stated that "more unexpected images" were presented in the New Idea mode (P4). Two participants summed this up by saying:

> [P20]: "…the [New Idea] mode gives me new ideas, the images and texts shown open my mind to new possibilities, not just what I already have in mind."

> [P31]: "…it was fun to see how many different variables came from the same sentence!"

In contrast, the narrow set of images generated from the Direct Input mode appeared to "confine [people's] thinking" (P3) and constrain the creative space in which people consider their ideas, which some participants found to be frustrating.

4.3.2 Contextual Preference for Direct Input Mode vs. New Idea Mode. The extent to which participants preferred images generated from the Direct Input mode as compared to the New Idea mode seemed to depend on the solidity of the idea that they were imagining. That is, if participants wanted to generate an image that was closely in line with what they were envisioning, they preferred the images generated from the Direct Input mode. In contrast, if participants did not have a clear idea of the image they desired or were less experienced in the field, they preferred the images generated from the New Idea mode. One participant summarized this idea, saying:

> [P4]: "I think it depends on the purpose. If it is purely creating a mood board, a diverse set of inspirational images would be helpful, helping me explore other ideas. If I am more clear on what I want to make, then a narrow set would be great."

These results suggest that both the New Idea and Direct Input modes serve uniquely important roles in the creative process, yet their utility may depend on the goal the user is aiming to achieve.

4.3.3 Idea Evolution. Participants tended to evolve their ideas over time in order to generate images that better satisfied their creative thinking. That is, the prompts would change with each iteration of image generation, which resulted in more satisfactory images being generated. For example, based on the output from their previous prompt, participants would edit and revise their prompt to more closely align with their intended image. As one participant noted:

> [P22]: "Each time I would input a new prompt, I would refine the text and be more specific and with that my results got better each time…the [ideas] that are generated were extremely helpful because it gave me better ideas to input."

Overall, participants' use of *DesignAID* allowed them to iterate through and expand upon ideas which contributed to a sense of satisfaction and inspiration during the creative process.

4.4 Summary of Pilots

In summary, participants seemed to be pleased with *DesignAID* in supporting their generation and selection of images in response to a prompt. Specifically, participants generally found the text prompts to be useful in inspiring them to generate and test new ideas. Furthermore, they preferred using the New Idea mode to expand their ways of thinking and generate images that they otherwise would not have imagined. In turn, participants tended to prefer the Direct Input mode when they had a clearer idea of the types of images to include in their mood boards. We draw from these preliminary qualitative findings to inform the design of a larger scale study with *DesignAID* while making user experience improvements.

5 QUANTITATIVE EVALUATION

After concluding our formative pilot studies, we performed a larger user study to answer the following research questions:

- How does interacting with image generation vs. image search influence designers' creative experience?
- How does the level of diversity in the prompts used for image generation and search affect the ideation process?

5.1 Methods

We employed a 2x2 mixed factorial design whereby participants were asked to create two mood boards depicting a "chair for children" [36]. Just as in the pilot studies, participants were asked to generate one mood board using the *high diversity* mode (New Ideas), whereby participants' input were fed into GPT-3 for prompt expansion. They also generated another mood board using the *low diversity* mode (Direct Input), whereby participants' search prompts were not expanded upon.

Furthermore, participants were randomly assigned to one of two conditions: *image generation* vs. *image search*. In the image generation condition, participants' prompts (after expansion by GPT-3) were fed into Stable Diffusion to produce images to include in their mood boards. As such, Stable Diffusion created images based on the participants' prompts. In the image search condition, we used Pinterest's algorithm to search for previously-created images using the description in the participants' input.

After creating each mood board, participants completed a questionnaire assessing their perceptions of their mood boards on various constructs such as inspiration and enjoyment (see below). Furthermore, after completing both mood boards, participants completed a final questionnaire assessing their preference for the high diversity vs. low diversity mood boards.

5.2 Participants

We recruited 102 participants from the online crowdsourcing site Prolific who indicated that they have experience or interest in the arts and/or design. Participants were given as much time as they needed to complete the study (M = 31.78 minutes, SD = 13.50 minutes) and were paid $15.00 for their participation. Only those who completed both mood boards and all the surveys were included in the analyses, resulting in a final sample size of 87 participants (28 male, 51 female, 7 non-binary or other, 1 not applicable; M_{age} = 25.51 years, SD_{age} = 5.73 years, $N_{Generation}$ = 45, N_{Search} = 42).

5.3 Measures

Participants completed five survey items measuring their subjective experience of using the *DesignAID* tool using a 1 (strongly disagree) to 7 (strongly agree) scale (see Table 1). Specifically, we measured subjective perceptions of inspiration [46], enjoyment [12], and usefulness [13]. Finally, we examined the amount of time it took for participants to create their mood boards (in minutes) and the number of image generation iterations they went through as behavioral indicators to better understand how participants used the *DesignAID* system.

Additionally, after participants had completed both the high and low diversity mood boards, they were asked to indicate which they thought was more inspirational, enjoyable, and useful. Participants could choose between high diversity, low diversity, or no preference.

5.4 Data Analysis

First, to examine whether the high and low diversity mood boards differed in terms of our outcome variables, we conducted several paired-sample t-tests. We then conducted several independent-sampled t-tests to examine whether our outcome variables differed between the image generation and image search conditions. Finally, to examine whether an interaction occurred between the two diversity and two mode conditions, we conducted several linear regression analyses with our outcome variables being regressed on both diversity and mode conditions. We used Cohen's D as a measure of effect size [9]. Importantly, because we had five outcome variables and two independent variables, we used Bonferroni corrections for all our p-values whereby we compared our significance against a p-value of .005 to reduce the incidence of Type 1 errors.

Item	Mean (Standard Deviation)
I experienced inspiration during this task	4.98 (1.61)
I liked doing the activity	5.26 (1.74)
Using this tool would enhance my effectiveness in my job	4.48 (1.86)
Time taken to complete mood board	9.36 (6.73)
Number of image generation iterations	10.80 (7.83)

Table 1. Survey Items and Behavioral Measures (survey scale: 1-7; time: minutes)

5.5 Results

High vs. Low Diversity

Participants tended to find the low diversity condition more enjoyable and useful than the high diversity condition (see Table 2a). Interestingly, these differences seemed to be more implicit, such that when participants were asked to directly indicate which mood board they found more enjoyable and useful, there was not a significant difference in the number of participants who indicated high diversity versus low diversity (see Table 2b). Furthermore, there were no significant differences between the high diversity and low diversity conditions in terms of inspiration.

When examining the behavioral measures, participants tended to spend more time in the high diversity condition than in the low diversity condition. However, there was no significant difference between the high diversity and low diversity conditions in terms of number of image iterations.

	Low Diversity		High Diversity		t (86)	p	Cohen's d
	M	SD	M	SD			
Inspiration	5.07	1.55	4.9	1.67	-0.94	0.35	-0.1
Enjoyment	5.63	1.39	4.89	1.97	-4.17	<.0001	-0.45
Usefulness	4.91	1.58	4.05	2.02	-4.62	<.0001	-0.5
Time	6.94	5.03	11.78	7.35	6.27	<.0001	0.67
Iterations	11.2	9.08	10.4	6.37	-0.92	0.36	-0.1

Table 2a. Paired-Sample T-Tests Examining the Differences Between High and Low Diversity Conditions

Item	Frequency			χ^2	df	p
	Low Diversity	High Diversity	No Preference			
Inspiration	31	25	31	0.83	2	0.66
Enjoyment	25	31	31	0.83	2	0.66
Usefulness	34	27	26	1.31	2	0.52

Table 2b. Chi-Square Goodness of Fit Test Examining the Differences Between High and Low Diversity Conditions

Image Generation vs. Image Search

When comparing the image generation vs. image search conditions, we found that participants tended to find the image generation condition more inspirational, enjoyable, and useful compared to those in the image search condition (see Figure 7; Table 3). Furthermore, those in the image generation condition tended to have fewer iterations of requests than those in the image search condition, although there was no significant difference in the amount of time participants took to make their mood boards between those in the image generation and image search conditions.

	Image Search		Image Generation		*t*	*p*	Cohen's *d*
	M	*SD*	*M*	*SD*			
Inspiration	4.44	1.73	5.49	1.31	-4.48	<.0001	-0.68
Enjoyment	4.43	1.88	6.03	1.16	-6.72	<.0001	-1.03
Usefulness	3.58	1.98	5.31	1.28	-6.79	<.0001	-1.04
Time	9.25	7.16	9.47	6.34	-0.21	0.84	-0.03
Iterations	13.8	9.77	8	3.72	5.1	<.0001	0.78

Table 3. Independent-Sample T-Tests Examining the Differences Between Image Generation and Image Search Conditions

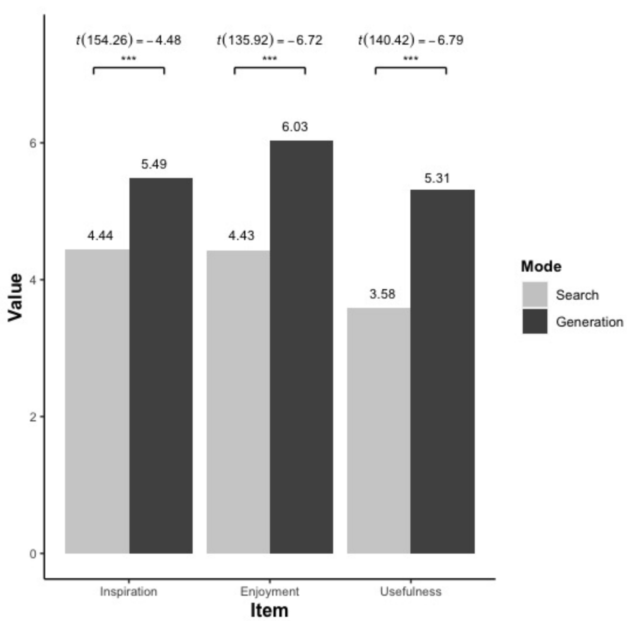

Figure 7: Differences in Inspiration, Enjoyment, and Usefulness between the Image Search and Image Generation Conditions $^{***}p$<.0001, $^{**}p$<.001, $^{*}p$<.005

Interactions Between Diversity and Mode

Interestingly, there were several significant interactions between diversity and mode. First, participants in the image generation condition rated the high diversity condition as being more inspirational but the differences were not significant for enjoyment and usefulness (see Figure 8, Table 4). However, for the image search condition, high diversity was significantly worse on all three measures.

There was marginal interaction between diversity and mode when predicting the amount of time participants took to complete the task. Specifically, it took them longer to complete their mood boards in the high diversity condition than the low diversity condition, although this difference was larger for those in the image search condition. Finally, there were no significant interactions between diversity and mode when predicting the number of iterations it took people to complete their mood boards.

	Coefficient	Standard error	p-value
Inspiration			
Intercept	3.88	0.23	<.0001
Diversity	1.12	0.32	0.0006
Mode	1.96	0.31	<.0001
Diversity*Mode	-1.83	0.44	<.0001
Enjoyment			
Intercept	3.57	0.22	<.0001
Diversity	1.72	0.31	<.0001
Mode	2.54	0.31	<.0001
Diversity*Mode	-1.87	0.44	<.0001
Usefulness			
Intercept	2.64	0.24	<.0001
Diversity	1.88	0.33	<.0001
Mode	2.71	0.33	<.0001
Diversity*Mode	-1.97	0.46	<.0001
Time			
Intercept	12.99	0.96	<.0001
Diversity	-7.47	1.35	<.0001
Mode	-2.33	1.33	0.08
Diversity*Mode	5.09	1.88	0.008
Number of Iterations			
Intercept	13.9	1.13	<.0001
Diversity	-0.21	1.6	0.89
Mode	-6.77	1.57	<.0001
Diversity*Mode	1.95	2.22	0.38

Table 4. Multiple Linear Regression Examining the Interactions Between Diversity and Mode

Overall, it appears that participants felt the image generation system (*DesignAID*) was more inspirational, enjoyable, and useful than the image search system (Pinterest) that we used as a baseline. However, we found that using a large language model (GPT-3) to generate highly diverse ideas had less apparent value. It was only marginally or not at all better on these dimensions for image generation and significantly worse for image search.

6 DISCUSSION

These results support the following claims: (1) Image generation systems can be more inspiring, enjoyable, and useful for designers than image search systems, (2) Increasing the diversity of ideas considered in this process has limited value. It provides some benefit for inspiration in image generation but otherwise has marginal or negative value, particularly for image search.

We view the empirical confirmation of the benefits of image generation systems as one contribution of this paper. But our results also raise an intriguing question: Why wasn't there more benefit from increasing the diversity of ideas considered?

Possible explanations include: Perhaps our system is not generating prompts with enough diversity. Maybe diverse ideas provide

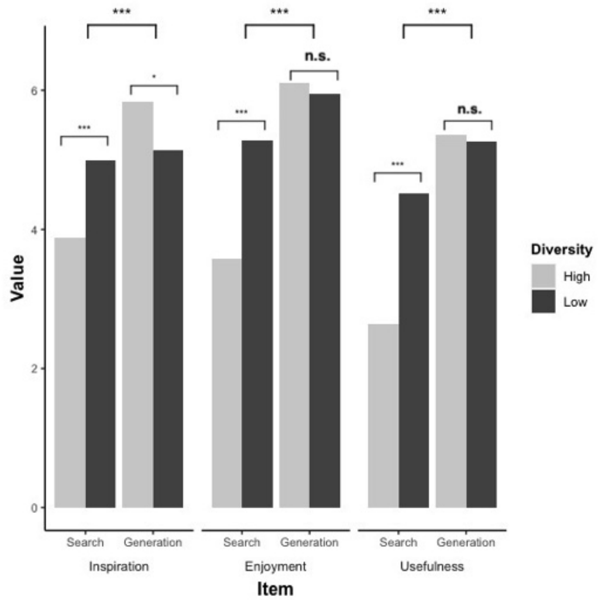

Figure 8: Interaction Between Diversity and Mode When Predicting Inspiration, Enjoyment, and Usefulness * *p*<.0001, ** *p*<.001, * *p*<.005**

more value in more open-ended tasks without prompts. Or, perhaps, just showing people images generated from diverse ideas isn't enough to overcome their strong design fixations.

Our results don't allow us to distinguish between these and other possible explanations, but we believe they suggest interesting questions for future research.

In any case, we view this work as a first step in distinguishing the role that diversity and generative AI could play in early-stage creative processes. In the following sections, we discuss implications for collective intelligence, improving this tool, and the potential broader impact of using generative AI in design.

6.1 Collective intelligence in human-computer systems

What kind of collective intelligence?

To study collective intelligence in any system, one can focus on either a group's *general* collective intelligence, that is, the group's ability to perform a wide range of different tasks [39, 50], or on its *specialized* collective intelligence, that is, its ability to perform a specific kind of task [30]. In this work, we focus on a group's specialized collective intelligence for performing the early stages of visual and physical design. And in this first study, we focus on the subjective experience of the designers.

Task allocation in human-computer systems

A key design question in any collective intelligence system is how to allocate tasks among group members [31], and in human-computer systems, the most critical aspect of this question is usually which tasks should be done by humans and which by computers.

For example, many designers have used computer drawing tools for years to capture and edit their designs and the human designers do most of the rest of the work. But we can expect the desirable task allocation in human-computer systems to continue changing as computer capabilities increase over time. In this study, we explore one example of this trend: the potential for new kinds of generative AI to expand the role computers play in the design process.

In particular, we focus here on a novel way of allocating more of the creative tasks of generating images to two kinds of generative AI: verbal (GPT) and visual (Stable Diffusion). The human designers in this case focus on specifying in English the kinds of images they want and then selecting the best ones from those generated by the AI systems.

We suspect that this general pattern of

- human specifies
- computer generates
- human selects

is likely to become a common type of task allocation for many uses of generative AI systems.

6.2 Improving *DesignAID*

Role of Diversity in Content Generation

Conceptual and aesthetic ideation are important and intertwined processes, but the dynamics of ideation along these two dimensions may differ throughout the creative process. Conceptual diversity in inspiration may be more important in the early stages of ideation, while aesthetic diversity may enable designers to more concretely explore visual variations of a concept. *DesignAID* provides a way for people to explore both conceptual or semantic diversity, facilitated by the New Idea mode, and visual diversity, facilitated by the generation of multiple visually distinct images.

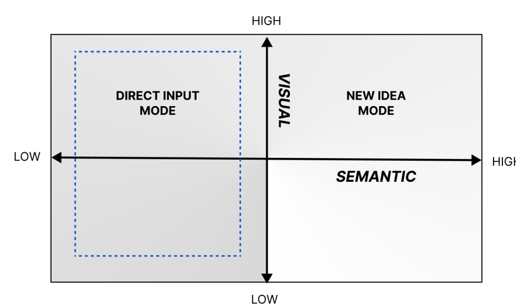

Figure 9: 2x2 diversity matrix with dimensions for semantic diversity and visual diversity.

While operating in the New Idea mode specifically, *DesignAID* supports wide exploration of the design space and what we consider divergent thinking both semantically and visually, as diagrammed in Figure 9. It is less obvious whether the Direct Input mode really supports convergent thinking or if it simply supports less divergent thinking, as the Direct Input mode can still include visual diversity but not semantic diversity. All user input to *DesignAID* is purely textual and the system does not use any knowledge of previous images in generating new ones. As such, the Direct Input mode will create multiple images with the same prompt, but those images could have varying degrees of differences between them and may actually continue to diverge rather than support convergence.

To better support truly convergent operation we believe it is important for *DesignAID* to also consider other information in the generation process. This could be important, for example, if a user wants an image to look very similar to another image with a specific modification, or to use the same palette or style of an input image when generating something new. Some prior work has demonstrated the ability to make granular updates to images by maintaining a mapping between input text and output image [19], and we believe this sort of operation should be considered for future systems supporting both convergent and divergent ideation.

Improving New Idea Generation with Prompt Engineering

In order to generate a diverse set of new ideas that can translate into effective visual images, it is important to understand how the text-generation and the image-generation elements work together. Recent research on diffusion-based image generation with tools such as DALL-E 2 and Midjourney has found that prompt engineering can be very important in guiding diffusion models in the right direction [28]. For example, specific keywords and phrases can often help generate output that is conceptually accurate and also aesthetically and stylistically desirable.

In our own exploration, we found that usage of stylistic adjectives and semantically similar nouns led to conceptually similar images with very unique stylistic interpretations. Instead of using the same word such as "chair" or "car" over and over, we found greater diversity and success with few-shot examples that used words with similar meanings like "couch", "loveseat", "sofa", "sportscar", "sedan", or "truck". Similarly using varying sentence structure led to more varied output. This informed how we designed the few-shot prompts we used with GPT-3.

While this few-shot approach worked for our early experiments, we believe that idea generation could be significantly improved with further prompt engineering, refining the few-shot examples given to the system, or using specifically curated fine-tuned models. These considerations are especially noteworthy when considering the body of emerging research aimed toward understanding the patterns and trends in how people are making practical use of diffusion-based image generation systems [49].

6.3 Importance of Augmenting Humans Rather than Replacing Them

We believe that systems like DesignAID have the potential to dramatically increase the creative output from human designers. Moreover, the DesignAID system was explicitly designed to *augment* these designers, not to *replace* them. In general, we believe that similar efforts to *augment* human creativity with computational tools are much more likely to be successful in the foreseeable future than attempts to completely *replace* human creativity with machines (e.g. work like that of [6]). Unless researchers and developers of generative AI tools focus on developing such synergistic human-computer systems, we–as a society–may fail to benefit from their potential. In a sense, therefore, the onus of responsibility for using computers in this way rests, not only on the users of such systems, but also on their creators.

7 CONCLUSIONS & FUTURE WORK

In this work, we developed *DesignAID*, a creativity support tool leveraging generative AI and semantic embeddings to facilitate broad design space exploration. We also investigated how designers used the tool to generate both diverse semantic ideas and visual representations of these ideas as they performed early-stage creative ideation. And we compared this form of image generation to a baseline of conventional image search.

We found that participants felt the image generation tool was more inspirational, enjoyable, and useful than the image search tool. But they felt that having automatically generated diverse ideas was worse on all these dimensions when using the image search tool and only somewhat better for inspiration when using the image generation tool.

We believe the results of our studies can help provide a foundation for better understanding how generative AI tools can increase the collective intelligence of human-computer teams doing creative design tasks. For example, we plan to develop future iterations of *DesignAID* to address the issues discussed above and to support groups of human users, not just individuals. We also anticipate running larger-scale quantitative evaluations of the tool and its use, including the quality of design outputs under different modes of operation.

In the long run, we hope that this work will help create "superintelligent" human-computer groups that can perform creative design and many other kinds of tasks more effectively than ever before.

ACKNOWLEDGMENTS

This research was supported by the Toyota Research Institute and also as part of the Collective Intelligence Mission of the MIT Quest for Intelligence.

REFERENCES

[1] Safinah Ali and Devi Parikh. 2021. Telling Creative Stories Using Generative Visual Aids. https://doi.org/10.48550/arXiv.2110.14810 arXiv:2110.14810 [cs].
[2] Abdullah Almaatouq, Joshua Becker, James P. Houghton, Nicolas Paton, Duncan J. Watts, and Mark E. Whiting. 2021. Empirica: a virtual lab for high-throughput macro-level experiments. *Behavior Research Methods* 53, 5 (Oct. 2021), 2158–2171. https://doi.org/10.3758/s13428-020-01535-9
[3] Oloff C. Biermann, Ning F. Ma, and Dongwook Yoon. 2022. From Tool to Companion: Storywriters Want AI Writers to Respect Their Personal Values and Writing Strategies. In *Designing Interactive Systems Conference (DIS '22)*. Association for Computing Machinery, New York, NY, USA, 1209–1227. https://doi.org/10.1145/3532106.3533506
[4] Oloff C Biermann, Ning F Ma, and Dongwook Yoon. 2022. From tool to companion: Storywriters want AI writers to respect their personal values and writing strategies. In *Designing Interactive Systems Conference*. 1209–1227.
[5] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, Vol. 33. Curran Associates, Inc., 1877–1901. https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html
[6] Andres Campero, Michelle Vaccaro, Jaeyoon Song, Haoran Wen, Abdullah Almaatouq, and Thomas W. Malone. 2022. A Test for Evaluating Performance in Human-Computer Systems. https://doi.org/10.48550/arXiv.2206.12390 arXiv:2206.12390 [cs].
[7] Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, et al. 2018. Universal sentence encoder. *arXiv preprint arXiv:1803.11175* (2018).

[8] Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St. John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, Brian Strope, and Ray Kurzweil. 2018. Universal Sentence Encoder for English. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Association for Computational Linguistics, Brussels, Belgium, 169–174. https://doi.org/10.18653/v1/D18-2029

[9] J Cohen. 1988. Statistical power analysis for the behavioral sciences. Lawrence Erlbaum Assoc., Hillsdale, NJ. *Statistical power analysis for the behavioral sciences. 2nd ed. Lawrence Erlbaum Assoc., Hillsdale, NJ.* (1988).

[10] British Design Council. 2005. The Design Process: The 'double diamond'design process model. *http://www. designcouncil. org. uk/about-design/how-designers-work/the-design-process/. Acesso em* 11, 12 (2005), 2013.

[11] Mihaly Csikszentmihalyi. 1997. *Creativity: Flow and the psychology of discovery and invention*. HarperCollins Publishers, New York, NY, US. Pages: viii, 456.

[12] Shayn Davidson. 2018. *A Multi-dimensional model of enjoyment: Development and validation of an enjoyment scale (ENJOY)*. Embry-Riddle Aeronautical University.

[13] Fred D Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly* (1989), 319–340.

[14] Nicholas Davis, Holger Winnemöller, Mira Dontcheva, and Ellen Yi-Luen Do. 2013. Toward a cognitive theory of creativity support. In *Proceedings of the 9th ACM Conference on Creativity & Cognition*. 13–22.

[15] Prafulla Dhariwal and Alexander Nichol. 2021. Diffusion Models Beat GANs on Image Synthesis. In *Advances in Neural Information Processing Systems*, Vol. 34. Curran Associates, Inc., 8780–8794. https://proceedings.neurips.cc/paper/2021/hash/49ad23d1ec9fa4bd8d77d02681df5cfa-Abstract.html

[16] Liane Gabora. 2002. Cognitive mechanisms underlying the creative process. In *Proceedings of the 4th conference on Creativity & cognition (C&C '02)*. Association for Computing Machinery, New York, NY, USA, 126–133. https://doi.org/10.1145/581710.581730

[17] Tim P German and H Clark Barrett. 2005. Functional fixedness in a technologically sparse culture. *Psychological science* 16, 1 (2005), 1–5.

[18] Tobin Hart. 1998. Inspiration: Exploring the Experience and its Meaning. *Journal of Humanistic Psychology* 38, 3 (July 1998), 7–35. https://doi.org/10.1177/00221678980383002 Publisher: SAGE Publications Inc.

[19] Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. 2022. Prompt-to-Prompt Image Editing with Cross Attention Control. https://doi.org/10.48550/arXiv.2208.01626 arXiv:2208.01626 [cs].

[20] Tanawat Hirunyawipada and Audhesh K Paswan. 2013. Effects of team cognition and constraint on new product ideation. *Journal of Business Research* 66, 11 (2013), 2332–2337.

[21] David G. Jansson and Steven M. Smith. 1991. Design fixation. *Design Studies* 12, 1 (Jan. 1991), 3–11. https://doi.org/10.1016/0142-694X(91)90003-F

[22] David G Jansson and Steven M Smith. 1991. Design fixation. *Design studies* 12, 1 (1991), 3–11.

[23] Youngseung Jeon, Seungwan Jin, Patrick C. Shih, and Kyungsik Han. 2021. FashionQ: An AI-Driven Creativity Support Tool for Facilitating Ideation in Fashion Design. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3411764.3445093

[24] Hyung-Kwon Ko, Subin An, Gwanmo Park, Seung Kwon Kim, Daesik Kim, Bohyoung Kim, Jaemin Jo, and Jinwook Seo. 2022. We-toon: A Communication Support System between Writers and Artists in Collaborative Webtoon Sketch Revision. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3526113.3545612

[25] Janin Koch, Nicolas Taffin, Andrés Lucero, and Wendy E. Mackay. 2020. SemanticCollage: Enriching Digital Mood Board Design with Semantic Labels. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference (DIS '20)*. Association for Computing Machinery, New York, NY, USA, 407–418. https://doi.org/10.1145/3357236.3395494

[26] Arthur Koestler. 1964. The Act of Creation. *British Journal for the Philosophy of Science* 16, 63 (1964), 255–257. Publisher: Taylor & Francis.

[27] 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. *Journal of Mechanical Design* 132, 4 (April 2010), 041003. https://doi.org/10.1115/1.4001110

[28] Vivian Liu and Lydia B Chilton. 2022. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–23. https://doi.org/10.1145/3491102.3501825

[29] Todd I Lubart. 2001. Models of the creative process: Past, present and future. *Creativity research journal* 13, 3-4 (2001), 295–308.

[30] Thomas W Malone. 2018. *Superminds: The surprising power of people and computers thinking together*. Little, Brown Spark.

[31] Thomas W Malone and Kevin Crowston. 1994. The interdisciplinary study of coordination. *ACM Computing Surveys (CSUR)* 26, 1 (1994), 87–119.

[32] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).

[33] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems* 26 (2013).

[34] Mohammad Amin Mozaffari, Xinyuan Zhang, Jinghui Cheng, and Jin L. C. Guo. 2022. GANSpiration: Balancing Targeted and Serendipitous Inspiration in User Interface Design with Style-Based Generative Adversarial Network. In *CHI Conference on Human Factors in Computing Systems*. 1–15. https://doi.org/10.1145/3491102.3517511 arXiv:2203.03827 [cs].

[35] Felix Müller-Wienbergen, Oliver Müller, Stefan Seidel, and Jörg Becker. 2011. Leaving the Beaten Tracks in Creative Work – A Design Theory for Systems that Support Convergent and Divergent Thinking. *Journal of the Association for Information Systems* 12, 11 (Nov. 2011). https://doi.org/10.17705/1jais.00280

[36] Jeffrey V Nickerson, Yasuaki Sakamoto, and Lixiu Yu. 2011. Structures for creativity: The crowdsourcing of design. In *CHI workshop on crowdsourcing and human computation*. 1–4.

[37] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125* (2022).

[38] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. Zero-Shot Text-to-Image Generation. In *Proceedings of the 38th International Conference on Machine Learning*. PMLR, 8821–8831. https://proceedings.mlr.press/v139/ramesh21a.html ISSN: 2640-3498.

[39] Christoph Riedl, Young Ji Kim, Pranav Gupta, Thomas W Malone, and Anita Williams Woolley. 2021. Quantifying collective intelligence in human groups. *Proceedings of the National Academy of Sciences* 118, 21 (2021), e2005737118.

[40] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis With Latent Diffusion Models. 10684–10695. https://openaccess.thecvf.com/content/CVPR2022/html/Rombach_High-Resolution_Image_Synthesis_With_Latent_Diffusion_Models_CVPR_2022_paper.html

[41] Pao Siangliulue, Kenneth C Arnold, Krzysztof Z Gajos, and Steven P Dow. 2015. Toward collaborative ideation at scale: Leveraging ideas from others to generate more creative and diverse ideas. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 937–945.

[42] Nikhil Singh, Guillermo Bernal, Daria Savchenko, and Elena L. Glassman. 2022. Where to Hide a Stolen Elephant: Leaps in Creative Writing with Multimodal Machine Intelligence. *ACM Transactions on Computer-Human Interaction* (Feb. 2022). https://doi.org/10.1145/3511599 Just Accepted.

[43] Steven M. Smith and Julie Linsey. 2011. A Three-Pronged Approach for Overcoming Design Fixation. *The Journal of Creative Behavior* 45, 2 (2011), 83–91. https://doi.org/10.1002/j.2162-6057.2011.tb01087.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.2162-6057.2011.tb01087.x.

[44] Zhu Sun, Jie Yang, Jie Zhang, Alessandro Bozzon, Long-Kai Huang, and Chi Xu. 2018. Recurrent knowledge graph embedding for effective recommendation. In *Proceedings of the 12th ACM conference on recommender systems*. 297–305.

[45] Duyu Tang, Furu Wei, Bing Qin, Nan Yang, Ting Liu, and Ming Zhou. 2015. Sentiment embeddings with applications to sentiment analysis. *IEEE transactions on knowledge and data Engineering* 28, 2 (2015), 496–509.

[46] Todd M Thrash and Andrew J Elliot. 2003. Inspiration as a psychological construct. *Journal of personality and social psychology* 84, 4 (2003), 871.

[47] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[48] Graham Wallas. 1926. *The art of thought*. Vol. 10. Harcourt, Brace.

[49] Zijie J. Wang, Evan Montoya, David Munechika, Haoyang Yang, Benjamin Hoover, and Duen Horng Chau. 2022. DiffusionDB: A Large-scale Prompt Gallery Dataset for Text-to-Image Generative Models. https://doi.org/10.48550/arXiv.2210.14896 arXiv:2210.14896 [cs].

[50] Anita Williams Woolley, Christopher F Chabris, Alex Pentland, Nada Hashmi, and Thomas W Malone. 2010. Evidence for a collective intelligence factor in the performance of human groups. *science* 330, 6004 (2010), 686–688.

[51] Tao Xu, Pengchuan Zhang, Qiuyuan Huang, Han Zhang, Zhe Gan, Xiaolei Huang, and Xiaodong He. 2018. Attngan: Fine-grained text to image generation with attentional generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1316–1324.

[52] Robert J. Youmans and Thomaz Arciszewski. 2014. Design fixation: Classifications and modern methods of prevention. *AI EDAM* 28, 2 (May 2014), 129–137. https://doi.org/10.1017/S0890060414000043 Publisher: Cambridge University Press.