Design Science

Prompting for products: investigating design space exploration strategies for text-to-image generative models

Leah Chong[1], I-Ping Lo[2], Jude Rayan[3], Steven Dow[3], Faez Ahmed[1] and Ioanna Lykourentzou[2]

[1]*Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA*
[2]*Department of Information and Computing Sciences, Utrecht University, Utrecht, Netherlands*
[3]*Department of Cognitive Science, University of California, San Diego, San Diego, CA, USA*

Abstract

Text-to-image models are enabling efficient design space exploration, rapidly generating images from text prompts. However, many generative AI tools are imperfect for product design applications as they are not built for the goals and requirements of product design. The unclear link between text input and image output further complicates their application. This work empirically investigates design space exploration strategies that can successfully yield product images that are feasible, novel and aesthetic – three common goals in product design. Specifically, users' actions within the global and local editing modes, including their time spent, prompt length, mono versus multi-criteria prompts, and goal orientation of prompts, are analyzed. Key findings reveal the pivotal role of mono versus multi-criteria and goal orientation of prompts in achieving specific design goals over time and prompt length. The study recommends prioritizing the use of multi-criteria prompts for feasibility and novelty during global editing while favoring mono-criteria prompts for aesthetics during local editing. Overall, this article underscores the nuanced relationship between the AI-driven text-to-image models and their effectiveness in product design, urging designers to carefully structure prompts during different editing modes to better meet the unique demands of product design.

Keywords: design space exploration, product design, prompt engineering, text-to-image generative AI

Received 19 December 2023
Revised 01 December 2024
Accepted 04 December 2024

Corresponding author
Ioanna Lykourentzou
i.lykourentzou@uu.nl

Des. Sci., vol. 11, e2
journals.cambridge.org/dsj
DOI: 10.1017/dsj.2024.51

1. Introduction

Rapid advancements in generative artificial intelligence (GenAI) have enabled the generation of novel and innovative content, such as texts and images, from simple text prompts. In product design applications, text-to-image models can produce images of designs from text prompts, enabling the exploration of multiple designs in shorter spans of time compared to the traditional method of manually rendering new designs. This functionality holds great potential for streamlining the iterative creative process in product design, particularly by facilitating design space exploration (DSE).

While text-to-image GenAI can enable the rapid exploration of diverse product design concepts, most existing tools are not engineered to account for the multi-faceted goals and requirements of product design, such as feasibility and aesthetics.

For example, current GenAI tools can generate a large number of designs, many of which are infeasible (Giannone et al., 2023; Regenwetter et al., 2023; Short, 2023). Chong and Yang presented a list of 16 different design objectives that are prevalent in design research and practice (Chong and Yang, 2023). This long – and yet non-exhaustive – list, underscores the complexity of parameters essential for designing a successful product. Unfortunately, the current GenAI possesses AI's inherent vagueness in the relationship between the input (i.e., communicated goal) and the generated output (i.e., images of designs), a property that renders GenAI tools insufficient for generating reliable product designs. For example, when the text prompt is "design of a mug that is ergonomic, sleek, and modern," it is unclear how the GenAI understands and maps the meaning of "ergonomic," "sleek" and "modern" onto the generated images. While one way to address this problem is to develop new models specifically trained for product design, current models possess a creative advantage, given the vast range of available image datasets compared to design-specific datasets like computer-aided design files. Therefore, this work aims to understand how off-the-shelf, promising yet imperfect text-to-image GenAI tools can be used to explore and refine product designs that are novel, aesthetically pleasing and feasible.

This work conducts a human subject experiment in which participants are asked to design bikes that are feasible, novel and aesthetic at the same time using Stable Diffusion 1.5 on an online platform called Leonardo.AI. At the time of running the experiment, Leonardo.AI was one of the few interactive tools that allowed the participants to easily use text-to-image generative models, such as various versions of Stable Diffusion, without the need to run custom Python scripts. The evaluations of the feasibility, novelty and aesthetics of the generated designs are collected via crowd-sourcing. The relationship between the participants' DSE strategies when using Stable Diffusion and the evaluation scores of their generated designs are analyzed. Key results from this study show that the goal orientation and the number of goals targeted by the prompts are more closely correlated with the design evaluation scores than the time users spend editing globally versus locally and the length of prompts. During early, broader exploration stages, multi-criteria prompts with feasibility and/or novelty goal orientation are found to be effective. Later in more refinement-focused stages, aesthetics-oriented prompts are suggested to be used.

The rest of this article is organized as follows. The next section provides a review of the background literature on DSE and text-to-image GenAI, identifying the research gap. Then, the purpose and the impact of this work are discussed, followed by the Method section, illustrating the study design, including descriptions of the task, participants and experimental procedure. Then, the results are presented and discussed, including a discussion of the limitations of this work. The paper concludes with a summary of the main findings and its implications for design research and practice.

2. Background

2.1 Design space exploration

DSE is a crucial step in the product design process, during which designers explore a wide range of potential designs (divergence) and then select and refine a fewer

https://doi.org/10.1017/dsj.2024.51 Published online by Cambridge University Press

selection of designs (convergence) (Cross, 2021; Georgiades et al., 2019). Divergence is important for successful design as it expands designers' creativity and increases the novelty and quality of their designs (Hilliges et al., 2007). It aims to prevent designers from limiting themselves to one or a few viable solutions too early by encouraging the formulation of a variety of potential solutions. Along with divergence, convergence is an equally important aspect of DSE. Once designers have explored enough, they must evaluate, select and refine the final solution(s) based on various design requirements, goals and preferences.

During DSE, designers engage in divergent thinking through various methods like problem reframing (Foster, 2021; Schon, 1987) and analogies (Mose et al., 2017). Generation and consideration of a large number of design options can be encouraged through the manipulation of a variety of characteristics, such as flexibility and imagery (Guilford, 1956). This process not only allows designers to maximize their creative output but also provides an opportunity to gain more insights into the problem and the design space. Prior research has attempted to find effective inspirations and methods to assist designers' divergent thinking using non-AI-based methods. For example, Thinklets (Briggs et al., 2001) is a creativity support tool that guides designers to think of ideas from multiple angles through open-ended questions. Ideation Decks (Golembewski and Selby, 2010) is another example, a set of cards that prompts designers to think about their design space from different angles. While these tools have shown some effectiveness, text-to-image GenAI presents the opportunity to increase the efficiency of divergence and significantly reduce designers' cognitive load by quickly generating multiple designs from text prompts. However, adopting this methodology means that designers must sacrifice some level of control in the design generation process (between text prompt and generated image).

Convergence is also a crucial aspect of DSE, during which designers make selections and/or mark preferences for certain aspects of generated designs (Mose et al., 2017). Often informed by data, designers choose specific design directions and make refinements to the designs as the most important dimensions of the problem space come into focus. Tools to support generative design exploration apply convergence methods in different ways. For example, the workflow in Dream Lens (Matejka et al., 2018) starts with the user defining the problem space. Then, the constraints and requirements from the resulting definition are used by the algorithm to generate a design. Additionally, GANCollage (Wan and Lu, 2023) updates its backend with the "user selection" every time the user requests "similar images" to accomplish the objective of choosing one final image. For effective design convergence, it is crucial to understand various user interactions during image exploration that could drive this process. Leonardo.AI, the tool used in this study, also includes various features that aim to facilitate design convergence, which will be explored in this work.

2.2 Image generative AI

With the recent advances in GenAI, there is great potential for these tools to effectively support creative processes. GenAI is a rapidly evolving field that involves the creation of algorithms and models capable of generating novel content in various domains, such as images, text and music. Its primary goal is to imitate the intricate creative process by leveraging existing datasets to identify underlying

https://doi.org/10.1017/dsj.2024.51 Published online by Cambridge University Press

patterns and yield outputs that closely resemble the characteristics of the training examples. Since 2020, discussions of various applications of GenAI, such as human resources, literature and art, have emerged. Specifically in product design, the potential of image GenAI as a tool for the early stages of the design process has been explored in some recent design literature (Lee and Chiu, 2023). There are primarily two types of models employed for image GenAI: Generative Adversarial Networks (GANs) and diffusion models.

GANs were introduced by Goodfellow in 2014 in the field of machine learning (ML). GANs are built using a pair of neural networks: a generator and a discriminator, operating on the principle that one network's gain is another network's loss. The generator is trained to generate new data samples, while the discriminator determines whether these samples are real or generated. Training continues until the discriminator's performance is above a certain threshold. Over the years, GANs have undergone significant refinement, incorporating methods such as injecting noise into the generator's input (Salimans et al., 2016), employing diverse loss functions (Mao et al., 2017) and applying regularization methods (Miyato et al., 2018), to promote the diversity of the generated data and improve the overall quality of the model outputs. With this refinement, the practical applications of GANs have expanded, serving as an effective model for image-to-image GenAI for DSE by rapidly creating a large number of possible designs (Chen and Ahmed, 2021; Li et al., 2021).

The diffusion model was introduced by Sohl-Dickstein et al. in 2015 as an alternative paradigm for GenAI (Nichol and Dhariwal, 2021). The diffusion model works by adding small random noise to the training data over multiple steps to produce a sequence of samples, then learning to recover the data by reversing this process. The performance of the model has been advanced continuously, giving rise to a flow-based generative model employing invertible transformations (Dinh et al., 2016), a continuous-time diffusion process called the Free-form Jacobian of Reversible Dynamics (FFJORD) model capable of generating high-quality samples with efficient inference (Grathwohl et al., 2018), and a new architecture that combined flow-based models with invertible 1×1 convolutions (Kingma and Dhariwal, 2018). Ho et al. (2019) further improved flow-based generative models, enhancing the quality and diversity of generated samples. Most of the current, widely-used text-to-image GenAI tools like DALL·E 2, Stable Diffusion and Midjourney are founded on diffusion models.

Diffusion models offer unique advantages compared to GANs. They guarantee more fine-grained control over the generated images, permit data quality and diversity manipulation (Giambi and Lisanti, 2023) and avoid mode collapse via a stable training process. A paper by OpenAI researchers (Dhariwal and Nichol, 2021) has indicated that diffusion models can achieve image sample quality superior to the GAN models. However, some main drawbacks of diffusion models are that they require longer training times and are computationally expensive because of the model's inherent complexity and the sequential nature of the diffusion process. This work uses Leonardo.AI, an online GenAI platform that provides an interface to work with various text-to-image diffusion models. Leonardo.AI is a particularly appropriate tool for this work as it offers global and local editing modes that are useful for investigating the users' DSE process. At the time of the experiment, Leonardo.AI was one of the few, if any, online text-to-image

https://doi.org/10.1017/dsj.2024.51 Published online by Cambridge University Press

GenAI platforms that allowed the users to seamlessly transition between global and local editing modes.

2.3 Research gap

Despite the promises of text-to-image GenAI in aiding the engineering design process, many current tools are imperfect tools for product design applications because of the various design goals and requirements, as well as AI's inherent vagueness in the relationship between the input (i.e., communicated goal) and the generated output. Therefore, there is an open question on how to use these promising yet imperfect GenAI tools for DSE and yield desirable designs. Only when this question is answered can people successfully utilize text-to-image GenAI for product design.

2.4 Research aims and significance

The purpose of this work is to close this gap in knowledge by investigating users' DSE strategies when using text-to-image GenAI and their impact on the generated outcome, specifically feasibility, novelty and aesthetics. We specifically observe the three design outcomes: feasibility, novelty and aesthetics because of their importance in DSE. The first two are common design goals during DSE based on the widely accepted definition that creative designs are both novel and appropriate (Hilliges et al., 2007; Miller et al., 2021; Amabile, 1996). Novelty is a fundamental goal in many design scenarios to ensure creativity and innovation (Mukherjee and Chang, 2023). Feasibility is a crucial goal particularly in product design, assuring that the generated designs align with real-world (often physical) constraints (Shah et al., 2003). At a conceptual stage like DSE, the assessment of feasibility is largely qualitative and estimated (Shah et al., 2003; Miller et al., 2021). The final design goal is aesthetics, which is also a significant component in product design as a major factor of market acceptance and product popularity (Bloch et al., 2003; Burnap et al., 2021; Lo et al., 2015).

The research question of this work is:

> *How do users' design space exploration strategies when using text-to-image generative AI influence the feasibility, novelty, and aesthetics of the generated product designs?*

Specifically, we want to understand whether and how the time users spend and the characteristics of prompts used have a significant impact on the feasibility, novelty and aesthetics of the generated outcomes. It is hypothesized that the more time spent exploring the design space without converging, the better the ratings of the generated outcome. Additionally, it is expected that the more prompting is focused on a goal, the more likely the rating for that goal will be higher. This work is expected to contribute to the design research community by suggesting what DSE strategies are effective when using GenAI for product design, specifically for designing feasible, novel and aesthetic products.

3. Method

A human-subject experiment is designed and conducted to examine how users leverage text-to-image GenAI to explore and create feasible, novel and aesthetic

https://doi.org/10.1017/dsj.2024.51 Published online by Cambridge University Press

designs. The feasibility, novelty and aesthetic ratings for generated designs are then collected via crowd-sourced evaluations. The experiment and evaluation data are analyzed to find DSE strategies that yield outcomes that successfully meet the design goals.

3.1 Human subject experiment

3.1.1 Experimental platform

The experiment is conducted using Stable Diffusion 1.5 on an online platform called Leonardo.AI. At the time of running the experiment, Stable Diffusion 1.5 was one of the most commonly used text-to-image GenAI, and Leonardo.AI was one of the few interactive tools that allowed the participants to easily use various versions of Stable Diffusion without the need to run custom Python scripts.

Leonardo.AI is an appropriate platform to study DSE strategies with, as it offers different editing modes, which we refer to in this work as global and local editing modes. Figure 1 displays what global and local editing modes entail in Leonardo.AI. The global editing mode primarily allows users to generate entire images by entering text prompts, while the local editing mode enables more detailed refinement of selected images using features like prompting, masking, and erasing. Given the functionalities, the use of the global and local editing modes is likely to correspond to the users' intention to diverge or converge design ideas respectively, which are crucial aspects of DSE.

An important thing to note is that given the experiment's goal to observe users' natural actions and behavior, this study does not provide any constraints to the users about how they should or should not interact with the different modes. Therefore, the correlation between the editing mode and divergence/convergence may not be definite as the users may intend to converge and refine in the global editing mode by entering more specific and detailed text prompts.

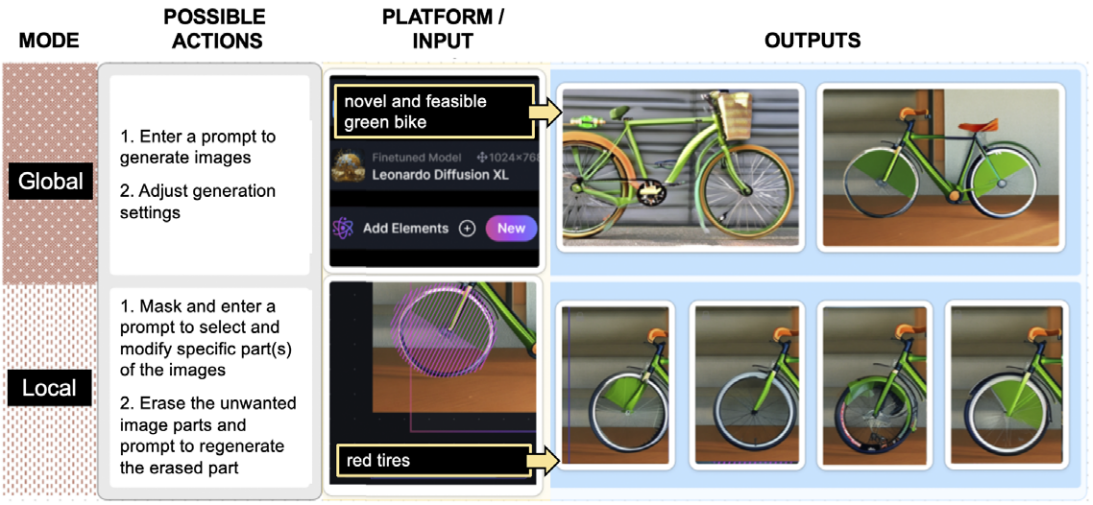

Figure 1. Global and local editing modes in Leonardo.AI and their example input and output. The global editing mode is primarily for generating entire images by entering text prompts, while the local editing mode is for more detailed refinement of selected images using features like prompting, masking and erasing.

https://doi.org/10.1017/dsj.2024.51 Published online by Cambridge University Press

3.1.2 Participants

A total of 15 participants are recruited and have completed the experiment. They range in age between 25 and 33 years and vary in their level of experience with GenAI tools. Regarding the level of experience with text-to-image GenAI tools, nine participants are first-time users, three are somewhat inexperienced, two are neither inexperienced nor experienced and one is somewhat experienced. The majority of the participants have a low level of prior experience with text-to-image GenAI tools because, at the time of the experiment, they were not yet prevalent. Six of the participants are male and nine are female. Their educational backgrounds range from bachelor's to doctorate degrees. Additionally, all participants demonstrate proficiency in English above level C1 and use English on a day-to-day basis.

A Google form for recruitment is employed to gather essential participant information, such as age, email address, education and level of experience with GenAI tools. It is also ensured that they viewed the tutorial video on Leonardo.AI and that their consent is obtained for participation and recording. Every participant is compensated with a 20 euro voucher for their participation in the experiment.

3.1.3 Experimental task

The participants are asked to create feasible, novel and aesthetically-pleasing bike design(s) using Stable Diffusion 1.5 on Leonardo.AI in 30 minutes. To provide a clear design context, the participants are given the "bike product designer" persona for a company named "22-Century Bike," as well as instructions to submit any number of bike designs. Throughout the experiment, they are encouraged to think aloud to collect data on their thoughts and decision-making processes. The following task description is provided to the participants:

> *You work for the company 22-century Bike as a bike product designer. Your job is to make one or more new bike designs. Your new bike design(s) should be feasible to manufacture (no triangle wheels!) and as unique/novel and aesthetically pleasing as possible. You can submit multiple designs.*
>
> *Remember to think aloud! The time will be about 30 mins.*

3.1.4 Procedure

At the start of the experiment, the participants are given a comprehensive pre-task tutorial. This tutorial aims to familiarize the participants with Leonardo.AI, explain the experimental task, and ensure that they possess adequate knowledge about bikes. First, to facilitate the participants' usage of Leonardo.AI's features, all features are demonstrated to them within the tool itself. Additionally, the participants are provided with a wiki page containing information about bike parts and types. Then, they are asked to complete a practice task on creating a vase for tulips using Leonardo.AI, giving the participants a chance to engage with the software and get their questions answered.

After the pre-task tutorial session, the main task begins. The participants are provided with the task description and are given 30 minutes to complete the task. Once the task is completed, the participants undergo a post-experiment interview and a brief questionnaire. The three interview questions are:

https://doi.org/10.1017/dsj.2024.51 Published online by Cambridge University Press

- Can you briefly describe your experience using Leonado.AI? What did you like about it? What did you dislike about it?
- Did you encounter any difficulties using the tool? If so, can you describe what they were?
- Are there any additional features or functionalities you would like to see added to the tool to improve?

These additional measures aim to gather qualitative data to understand the quantitative results further. Then, the questions in the post-experiment questionnaire ask the participants to report how easy it was to use each feature of the tool, how important they think each feature of the tool is, and how feasible, novel and aesthetic they think their final designs are. These questions are answered in a 5-point Likert scale.

3.1.5 Crowd-sourced design evaluations

The final set of 18 images is submitted by the 15 participants (one image each by 12 participants and two images each by three). Crowd-sourced evaluations are conducted using Google Forms to assess the feasibility, novelty and aesthetics of these designs. The evaluation questionnaire asked raters to evaluate each bike image on its feasibility, novelty, and aesthetics based on a 5-point Likert scale ranging from "Strongly Disagree" to "Strongly Agree." Some example questions are shown in Figure 2.

Raters are recruited from the Prolific platform, which is a platform for researchers to recruit and manage participants from a large, global participant pool for online studies. Prolific is selected as the crowd-sourcing platform as it ensures high-quality responses through a more rigorously curated participant pool compared to other commonly-used platforms, such as Amazon Mechanical Turk (Douglas et al., 2023; Peer et al., 2022). In total, 10 raters evaluated the images, allowing for a broader range of perspectives and opinions to be considered.

4. Results

In this work, DSE strategies are observed by the participant's actions in the global and local editing modes in Leonardo.AI. Within each editing mode, the participants' actions are mostly done via text prompting. Therefore, this work examines the participants' prompting action characteristics to gain insight into their exploration strategies. Along with the participants' exploration strategies, this section also presents the correlation results between these strategies and the feasibility, novelty and aesthetics ratings of the generated outcomes.

Three major action characteristics in the global and local editing modes are examined: length, goal orientation and multi versus mono-criteria of the prompt. These three characteristics have been selected because they are frequently discussed in the prompt engineering literature. For instance, Xie et al. (2023) performed a log analysis and found a correlation between the length of the prompt and the quality of the generated image. From a prompt construction perspective, studies have discussed mono-criteria prompts having higher accuracy than multi-criteria prompts (Tan et al., 2020; Wei et al., 2022), meaning that targeting multiple design objectives in a single prompt might not accurately express the designer's

https://doi.org/10.1017/dsj.2024.51 Published online by Cambridge University Press

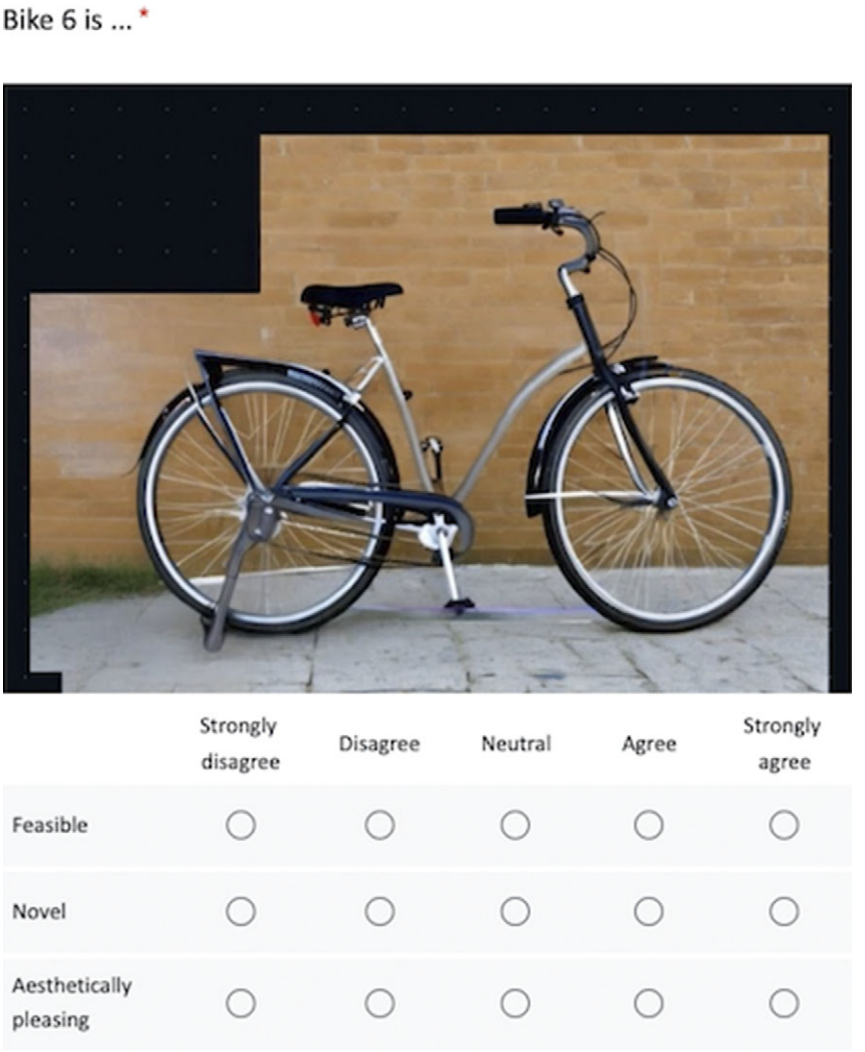

Figure 2. Example problem in the crowd-sourced image evaluations.

intentions. Finally, PromptMagician is a tool that suggests keywords to be added to the text prompt to enhance the alignment of the generated images with the intended vision of the creator (Feng et al., 2023).

It is important to note that three out of the 15 participants submitted two bike designs and none more than two. For these three participants, the average data of their two bike creation processes are used throughout the analyses. Therefore, in this work, one data point consistently represents each participant.

Before conducting each statistical analysis presented in this section, the normality of the included data is tested via Shapiro–Wilk test. If the data are normal, a two-sample t-test and Pearson's correlation test are conducted for two sample comparisons and correlation analyses respectively. If the data are not normal, the Wilcoxon signed-rank test and Spearman's rho test are conducted.

https://doi.org/10.1017/dsj.2024.51 Published online by Cambridge University Press

4.1 Global versus local editing

One notable aspect of users' exploration strategy with respect to Leonardo.AI is how they choose to allocate their time between the global and local editing modes. As shown in Figure 1, the global and local editing modes are available on Leonardo.AI for the participants to either change the entirety of the AI-generated image using text prompts or to change only a part of this image using various features like masking and erasing, respectively. The visualization of each participant's time distribution among global and local editing modes is shown in Figure 3.

Most participants (14 out of 15) engage in both global and local editing, while only one user exclusively focused on global editing. It is observed that participants all start in the global editing mode, and after 4.0 prompts in the global editing mode on average, are inclined to select an image for local refinement. Furthermore, it was observed that the participants who entered the local editing mode never reverted back to the global editing mode, unless they were starting a new design.

On average, the participants allocate about 38.5% of their time to global editing and concentrating on general modifications, and the remaining time (61.5%) to local editing and refining specific details to meet the design goals. To investigate the split between global and local editing modes further, the number of prompts the participants enter in the two modes is observed. About 36.7% of the prompts (on average 4.0 prompts) are entered in the global editing mode and the rest (63.3%) in the local mode, a fairly consistent result with the time split result. Overall, the participants spend more time in the local editing mode than in the global editing mode, though without statistical significance (Wilcoxon signed-rank test, $p = 0.2$ for time and $p = 0.1$ for number of prompts).

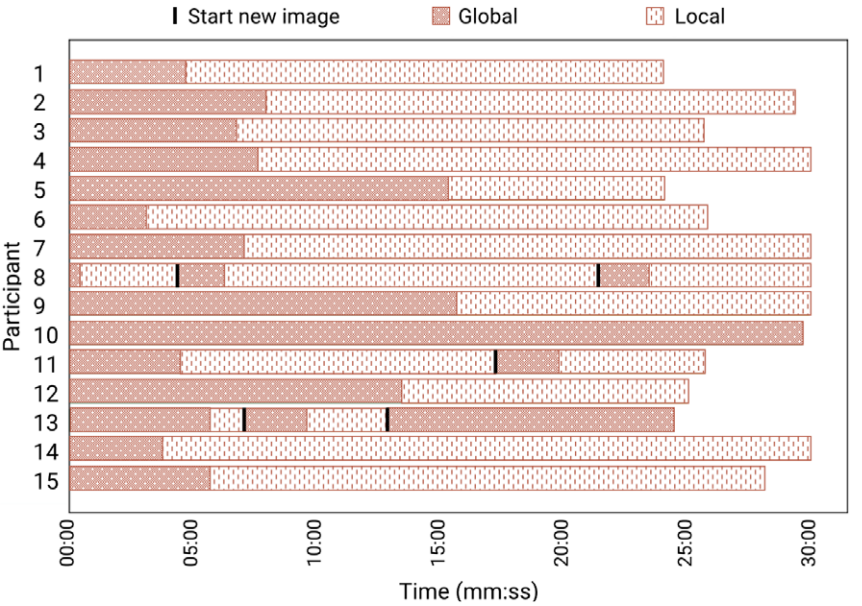

Figure 3. The distribution of time spent in global and local editing modes. On average, the participants spent about 38.5% of their time in the global editing mode, and the other 61.5% in the local editing mode.

https://doi.org/10.1017/dsj.2024.51 Published online by Cambridge University Press

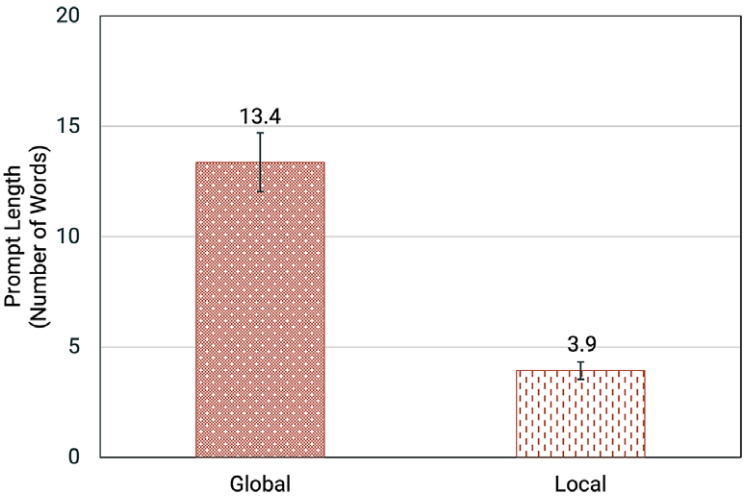

Figure 4. Average Prompt Length. The participants use significantly longer prompts in the global editing mode compared to the local editing mode.

4.2 Prompt length

The participants' actions via prompting in the global and local editing modes are explored to extract their exploration strategies. The first action characteristic is the length of prompts. The average length of prompts are demonstrated in Figure 4. On average, the participants' prompts are about 7.7 words long, and the prompts they use in the global editing mode (13.4 words long on average) tend to be longer than those in the local editing mode (3.9 words long on average) (t-test, p = 1.1e-3). The length of the longest prompt is 47 words and the shortest has only one word.

4.3 Mono versus multi-criteria prompts

Given that the participants are asked to design for three different goals simultaneously, it is important to examine how they orient their prompts for the goals in the global and local editing modes. For this purpose, every prompt by the participants is labeled by the authors as feasibility, novelty and/or aesthetics-oriented. Prior to labeling, the three authors agree on the descriptions of the three orientations of prompts shown in Table 1, which are used to guide the labeling process.

The authors label the prompts separately according to the descriptions in Table 1, and the final labels are determined based on the majority agreement. Multi-criteria labeling is allowed. For example, if a prompt is relevant to both feasibility and novelty, it will be labeled as a both feasibility and novelty-oriented prompt. When there are negative prompts (prompts to eliminate components from an image), they are considered together with the main prompts. For instance, if the main prompt is primarily feasibility-oriented and the negative prompt is aesthetics-oriented, these prompts are considered as a single prompt that is both feasibility and aesthetics-oriented. Example prompts with different goal orientation(s) are shown in Table A1.

Many participants employ prompts that are oriented towards multiple design goals (i.e., multi-criteria prompts), such as "square tire bike with bottle cage and red

https://doi.org/10.1017/dsj.2024.51 Published online by Cambridge University Press

Table 1. Descriptions of feasibility, novelty and aesthetics-oriented prompts

Goal orientation	Description
Feasibility	Contains explicit words related to manufacturing, addition or modifications of bike parts, Is related to the usability (whether it successfully functions/can be used) and/or manufacturability
Novelty	Contains explicit words related to novelty or uniqueness, prompting for any bike parts that are not in traditional bikes
Aesthetics	Contains any words related to the look/visual feel/dimensions of the bike

chain ring" which is both novelty and aesthetics-oriented, while others concentrated on prompting for a single goal (i.e., mono-criteria prompts), such as "a bike with special design" which is only novelty-oriented. Therefore, the average percentage split of multi- versus mono-criteria prompts in the global and local editing modes are observed to understand the participants' exploration strategy. Figure 5 shows the results.

Overall, 52.0% of the prompts are mono-criteria and the rest (48.0%) are multi-criteria, therefore showing a relatively equal split (t-test, $p = 0.7$). This overall split is most likely because of the contrasting split in the global and local editing modes. In the global editing mode, there are more though not statistically

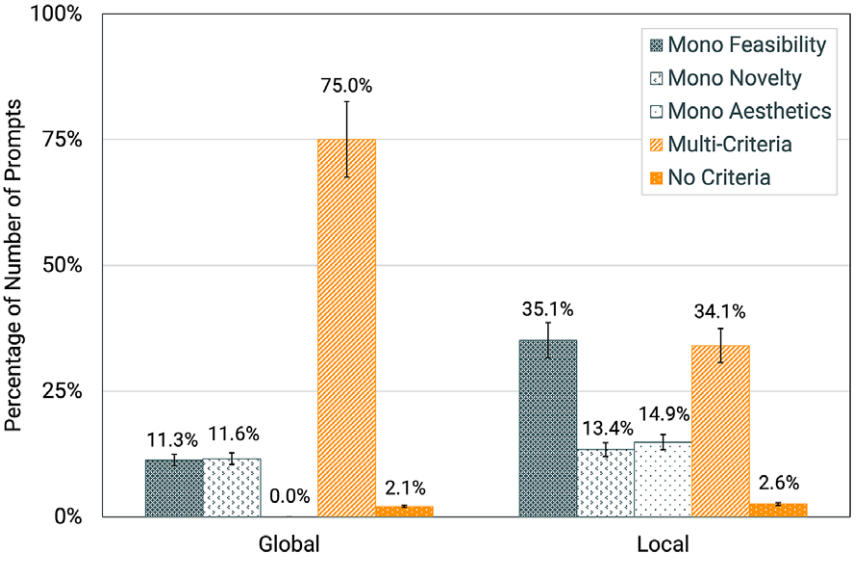

Figure 5. Percentage of mono-criteria (feasibility, novelty and aesthetics-oriented), multi-criteria and no-criteria prompts. the participants use much more multi-criteria prompts in the global editing mode than in the local editing mode. Most of the mono-criteria prompts in the global editing mode are feasibility or novelty-oriented, while those in the local editing mode are mostly feasibility-oriented with some being novelty and aesthetics-oriented.

https://doi.org/10.1017/dsj.2024.51 Published online by Cambridge University Press

significant, multi-criteria prompts (75.0%) than mono-criteria prompts (22.9% = 11.3% + 11.6% + 0.0%) (Wilcoxon signed-rank test, p = 0.07), while in the local editing mode, there are more, though not statistically significant, percentage of mono-criteria prompts (63.4% = 35.1% + 13.4% + 14.9%) (Wilcoxon signed-rank test, p = 0.1) than multi-criteria prompts (34.1%). Therefore, the results show a tendency among the participants to tackle multiple design goals at once in the global editing mode, while taking one goal at at time in the local editing mode.

4.4 Goal orientation of prompts

Figure 6 shows the percentage split of feasibility, novelty and aesthetics-oriented prompts. It is important to note that both mono and multi-criteria prompts are considered in these results; for example, the percentage of feasibility-oriented prompts includes the mono-criteria ones that are only feasibility-oriented, as well as the multi-criteria ones that target feasibility along with other goals.

In the global editing mode, all three goals are targeted often in the participants' prompts without any statistical difference between the percentages of these orientations (78.5% feasibility-oriented, 73.8% novelty-oriented and 68.7% aesthetics-oriented). In contrast, in the local editing mode, the participants used a much higher percentage of feasibility-oriented prompts (59.4%) than novelty-oriented prompts (19.8%) (Wilcoxon signed-rank test, p = 0.02). Comparing the results in the global and local editing modes, it is observed that the prompts in the global editing mode often target all three goals, while the prompts in the local editing mode are more likely to target feasibility over the other two goals.

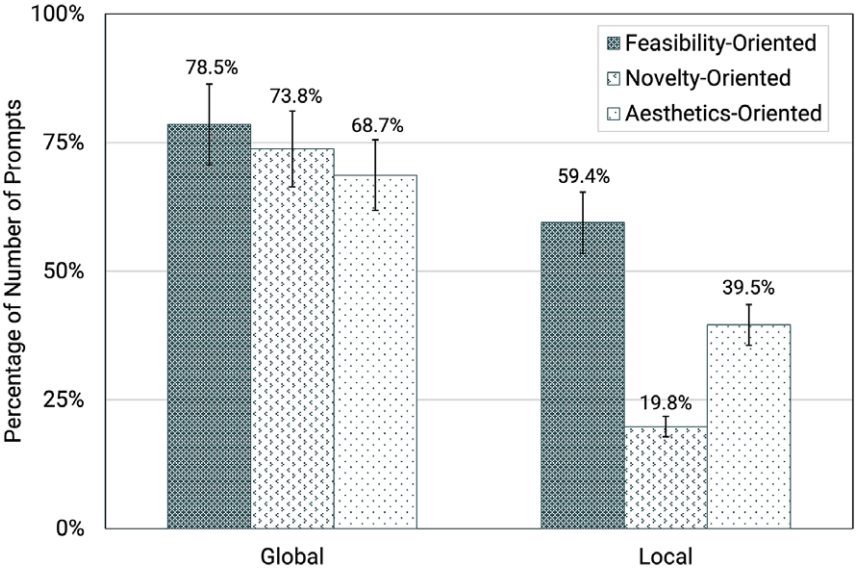

Figure 6. Percentage of Feasibility, Novelty and Aesthetics-Oriented Prompts. The results shown here include both mono and multi-criteria prompts, therefore not adding up to 100%. While targeting all three goals often in the global editing mode, the participants tend to focus more on feasibility and aesthetics than novelty in the local editing mode.

https://doi.org/10.1017/dsj.2024.51 Published online by Cambridge University Press

4.5 Correlation with the design outcome

Correlation analyses between the observations above and the feasibility, novelty and aesthetic ratings of the generated images are conducted to identify the exploration strategies that yield desirable design outcomes. The ratings are determined by the crowd-sourced evaluations described in the Method section.

4.6 Global versus local editing and prompt length

The correlations between the time the participants spent and their prompt length in the global and local editing modes with the design outcome are examined. No statistically significant correlations are found, meaning that neither how much time is spent nor how many prompts are used in the global versus local editing modes are related to any ratings of the design outcome.

4.7 Mono versus multi-criteria and goal orientation of prompts

The participants' exploration strategy is also studied via two of their prompting characteristics in the global and local editing modes: mono versus multi-criteria and goal orientation of prompts. The correlations between these characteristics and the feasibility, novelty and aesthetic ratings of the bikes are computed, as demonstrated in Figure 7. The correlation graphs of the statistically significant results are included in the Appendix in Figure A1.

During the global editing mode, the percentage of feasibility-oriented prompts is significantly correlated to the feasibility rating of the generated image (Spearman's rho test, rho = 0.6, p = 0.02). This means that using more feasibility-oriented prompts during the global editing mode can significantly boost feasibility ratings. Such correlations are not shown between novelty and aesthetics-orientated prompts and their corresponding ratings (Spearman's rho test, rho = 0.3, p = 0.3 for both).

Interestingly, however, the percentage of novelty-oriented prompts in the global editing mode is negatively correlated to the aesthetics rating (Spearman's rho test, rho = −0.6, p = 0.02). When more novelty-oriented prompts are used in the global editing mode, the generated design is less aesthetically-pleasing.

Finally, the percentage split between mono and multi-criteria prompts in the global editing mode is found to be negatively correlated to the feasibility rating (Spearman's rho test, rho = −0.5, p = 0.03). More multi-criteria prompts and fewer mono-criteria prompts can help increase the feasibility rating of the design outcome.

In the local editing mode, the only significant correlation is found between the percentage of aesthetics-oriented prompts and the feasibility rating of the outcome. This correlation is positive, meaning that the more aesthetics-oriented prompts are used in the local editing mode, the higher the feasibility rating of the generated image is (Spearman's rho test, rho = 0.6, p = 0.02).

The final result to note is the general positive relationship between feasibility and aesthetics, while they both show a negative relationship with novelty. The results regarding goal orientation of prompts in Figure 7 mostly demonstrate positive correlation coefficients between feasibility and aesthetics. Novelty has negative correlation coefficients with feasibility and aesthetics. This result is supported by the positive correlation between the percentage of feasibility-oriented and aesthetics-oriented prompts found in the Goal Orientation of Prompts section.

https://doi.org/10.1017/dsj.2024.51 Published online by Cambridge University Press

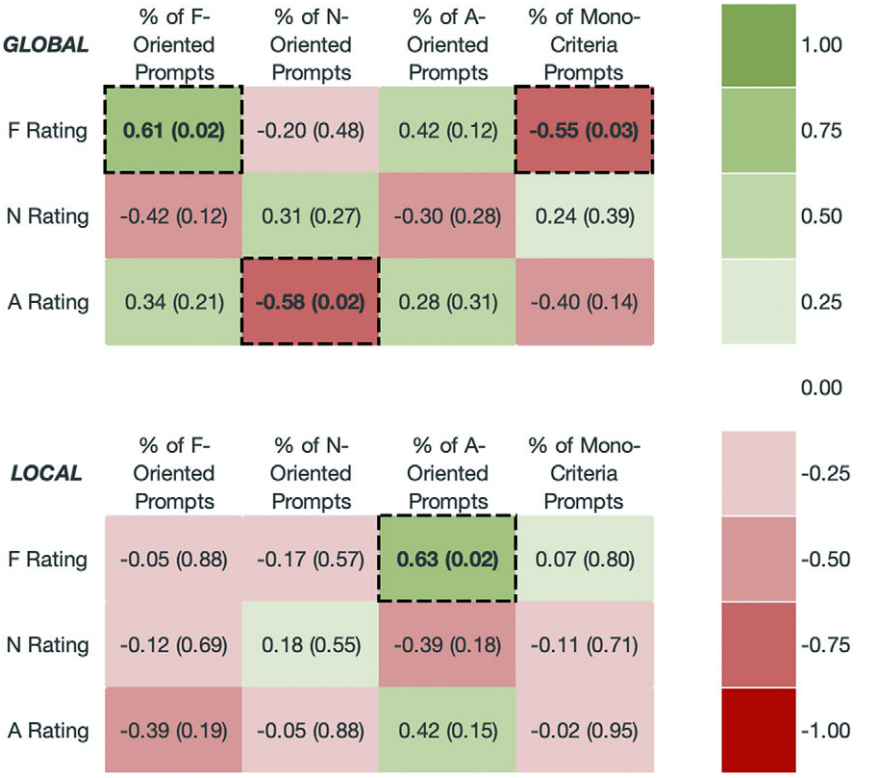

Figure 7. Correlation between the prompting characteristics (mono versus multi and goal orientation) in global and local editing modes and the crowd-sourced design ratings (feasibility, novelty and aesthetics). The values are indicated as (correlation coefficient) (p-value). The results bolded with dashed lines are statistically significant at 5%.

5. Discussion

The purpose of this work is to address the research question:

> *How do users' design exploration strategies when using image generation AI tools influence the feasibility, novelty, and aesthetics of the generated outcomes?*

This section discusses the answers to this question with the results found in this work, as well as their implications for design research and the use of text-to-image GenAI in product design. Then, the limitations of this work and the areas for future research are discussed.

As the discovered insights are discussed, it is crucial to keep in mind that they display the DSE strategies and outcomes of novice users of text-to-image GenAI tools. This is because this work was conducted as an exploratory study before text-to-image GenAI became prevalent. Even with the comprehensive tutorial at the start of the experiment, novice users may experience a learning curve. Their lack of experience may also limit their ability to make full use of the tool's features and affect the way they interact with it. Some of our qualitative analysis results show

https://doi.org/10.1017/dsj.2024.51 Published online by Cambridge University Press

that in fact, the majority of the participants only used about half the features the tool offered.

The results in this work first show that people commonly employ a combination of global and local editing with a significant focus on local refinements, spending approximately 61.5% of their time locally editing the images. Once in the local editing mode, they do not return to the global editing mode. This behavior may be because of the complexity of the steps on Leonardo.AI to return to the global editing mode, especially given the time constraint. It could also be a reflection of "design fixation" (Jansson and Smith, 1991; Linsey et al., 2010), which describes people's tendency to overly focus on their initial ideas without considering diverse set of ideas.

Despite observing similar human behaviors from prior works, this work demonstrates no relationship between the time spent editing the image globally and locally, and the feasibility, novelty and aesthetics ratings of the final design. At a glance, this is surprising in light of some prior findings about the positive impact of more time spent exploring the design space in the design process before converging on a solution (Kudrowitz and Dippo, 2013). However, it is important to note that the global and editing modes in this study may not be the exact reflections of divergence and convergence of ideas. Considering the features of text-to-image GenAI tools, such as Leonardo.AI, the users can easily fixate on a design even in the global editing mode by inserting conceptually similar prompts rather than exploring different design options with conceptually far prompts. Therefore, the most precise way to interpret the result is that the feasibility, novelty and aesthetics ratings of the final design are not related to the amount of time people spend on exploring and editing the overall image versus specific parts of the image. This means that it does not matter how much time is spent on global versus local editing to generate a highly-rated design.

Secondly, the prompts in the global editing mode tend to be longer, multi-criteria and more heavily goal-oriented than those in the local editing mode. These findings make sense based on the nature of the global editing mode, which is to edit the entire image rather than focusing on specific parts. This seems to lead people to prompt with more words and goals all at once. In the global editing mode, it is found to be beneficial to use more feasibility-oriented prompts and less novelty-related prompts for better feasibility and aesthetics ratings respectively. Also, it is good to keep using multi-criteria prompts over mono-criteria prompts as this helps to reach a higher feasibility rating.

In contrast, the prompts in the local editing mode tend to be short and single goal oriented (often feasibility rather than novelty or aesthetics). Interestingly, the correlation results show that it is good to use more aesthetics-oriented prompts in the local editing mode as this increases the feasibility rating of the final design. Consistent with this result, the close relationship between feasibility and aesthetics is discovered throughout the results. Specifically, the percentage of feasibility-oriented prompts demonstrates a positive correlation to the percentage of aesthetics-oriented prompts. This implies that people intuitively combine prompts that improve both feasibility and aesthetics. Furthermore, though not statistically significant, aesthetics-oriented prompts lead to designs that not only satisfy the raters' visual preferences, but also are rated feasible. This synergistic relationship between aesthetics and feasibility is often referred to as the "aesthetic-usability" or "aesthetic-utility" effect (Kurosu and Kashimura, 1995; Sonderegger and Sauer,

https://doi.org/10.1017/dsj.2024.51 Published online by Cambridge University Press

2010), which underscores the phenomenon where people perceive aesthetically pleasing designs as more usable and effective, even when the functionality remains unchanged.

Finally, the results in this work do not demonstrate a clear way to increase novelty ratings of the final product designs. This suggests that increasing novelty may require other prompting strategies besides manipulating the number and type of goal orientations. For example, Ma et al. (2023) suggest leveraging few-shot learning. Moreover, the results in this work consistently display that novelty has a marginally negative relationship with both feasibility and aesthetics, agreeing with prior works that discovered a tradeoff between feasibility and novelty of designs (Mukherjee and Chang, 2023). The relationship between novelty and aesthetics seems to be more nuanced, as there are more variance in its findings, such as Hung and Chen showing an inverted-U relationship (Hung and Chen, 2012).

In summary, when using text-to-image GenAI tools for product design, it is critical to pay attention to the number and type of goals that are targeted in the prompts perhaps more than to the time spent editing globally and locally and the length of prompts. While longer prompts may seem comprehensive and therefore always effective, they do not consistently lead to better feasibility, novelty and aesthetic results. Therefore, rather than simply aiming for long prompts, users are advised to strategize to employ specifically oriented prompts to guide DSE more effectively. Incorporating targeted text within prompts can be advantageous in achieving the desired outcomes. What type of targeted prompts should be used then? Initially in the global editing mode, multi-criteria prompts that are both feasibility and aesthetics-oriented are suggested to be used, rather than mono-criteria or novelty-oriented prompts. Then in the local editing mode, it is good to use many aesthetics-oriented prompts. To highlight, aesthetics-oriented prompts are recommended to be employed throughout the DSE process as they benefit the overall rating of the design, especially the feasibility rating.

6. Limitations and future work

There are several limitations to this work. First, the study is conducted on a single platform (Leonardo.AI) and with a text-to-image model (Stable Diffusion 1.5), which may limit the generalizability of the findings. Depending on the performance of the model, especially as a result of the training dataset and/or fine-tuning, the same text prompts may produce different outcomes. However, the core concept of this paper about using text-to-image models for global and local editing during DSE is generalizable as this is a fundamental tradeoff in design. Therefore, the DSE techniques suggested in this work may be helpful across different models though they may be adopted with caution. Furthermore, the insights in this work about users' DSE behavior using GenAI beyond the specific strategies could have longer-lasting value since human cognitive abilities are relatively stable compared to the dynamic computational tools. To further explore the generalizability, replicating the experiment with other text-to-image models and platforms can bolster the findings of this work.

Secondly, the experiment has a small sample size, which may lead to unreliable statistical results. Considering this limitation, besides the results that demonstrate a very strong statistical significance, we only analyze and discuss the results qualitatively. It is also critical to note that this is an exploratory study to gain preliminary

https://doi.org/10.1017/dsj.2024.51 Published online by Cambridge University Press

insights that must be confirmed further through future studies with a larger sample size.

Another limitation is that some individual biases may be present in the bike ratings. We chose to collect the ratings from 10 raters and only utilize the aggregate values for the analysis. While this method resolves the individual biases more than other methods, such as single expert ratings, the potential discrepancies in the raters like their understanding of feasibility, novelty and aesthetics may affect the results. Exploring the rater differences in future works can further clarify the results and insights in this work.

Furthermore, the potential impact of users' prior experience with GenAI tools and background in design on the outcomes is not directly studied. This is because, at the time of conducting the experiment, text-to-image GenAI tools were not yet widely used, leading to most of the participants in our study being novice users. Incorporating user profiles, such as experience level in design or familiarity with GenAI tools, could provide deeper insights into how different users might benefit from different strategies. Another interesting approach would be to study how the users' prolonged use of GenAI tools impacts their DSE strategies.

In addition, the results of this study, especially those related to time, depend on the usability of the GenAI tool. For example, if the tool's functionalities in the local editing mode are difficult to navigate around, users may spend much more time in the local editing mode than if the functionalities are easier to use. Therefore, future work can explore how improving the usability of GenAI tools affects the users' actions, DSE strategies and design outcomes.

Lastly, it is important to clarify the scope of these findings that it only applies in the context of using GenAI for DSE. If the design is intended for physical manufacturing and beyond, it is crucial to consider additional tools for 3D modeling or physics-based evaluations, and the results from this work do not directly apply.

7. Conclusions

This work investigates how DSE can be conducted effectively when working with GenAI tools, which are often not catered to product design applications. Given the increasing popularity of these tools, many product designers will likely attempt to use these tools in their DSE process. Given the findings in this work, they are likely to be able to more successfully achieve their design goals, particularly feasibility, novelty and aesthetics, by strategically guiding their interactions with the tool based on this work rather than using it merely based on their intuition.

The findings in this work provide valuable insights into the user approaches and strategies for DSE success when using text-to-image GenAI tools. By considering these insights and recommendations, researchers and designers can enhance the effectiveness and usability of such tools, and users can utilize these strategies to pursue better product design outcomes. For instance, researchers of Leonardo.AI can study the accurate relationship between feasibility-oriented prompts and the feasibility of the output design during the global editing mode. Additionally, designers can more effectively utilize GenAI for product design, using multi-criteria, feasibility and aesthetics-oriented prompts in the early stages, then mono-criteria, aesthetics-oriented prompts in the later stages.

https://doi.org/10.1017/dsj.2024.51 Published online by Cambridge University Press

Financial support

Prof. Ahmed and Prof. Lykourentzou extend their gratitude to the MIT MISTI Netherlands Program for their funding support. Prof. Dow thanks the National Science Foundation for research funding under NSF grant #2009003.

References

Amabile, T. 1996 *Creativity in Context*. Westview Press.

Bloch, P. H., **Brunel, F. F.** & **Arnold, T. J.** 2003 Individual differences in the centrality of visual product aesthetics: Concept and measurement. *Journal of Consumer Research* **29**(4), 551–565.

Briggs, R., **De Vreede, G.**, **Nunamaker, J.** & **Tobey, D.** 2001 ThinkLets: Achieving predictable, repeatable patterns of group interaction with group support systems (GSS). In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*, 9pp.

Burnap, A., **Hauser, J. R.** & **Timoshenko, A.** 2021 Design and evaluation of product aesthetics: a human–machine hybrid approach. SSRN: 3421771.

Chen, W. & **Ahmed, F.** 2021 PaDGAN: Learning to generate high-quality novel designs. *ASME Journal of Mechanical Design* **143**(3), 031703.

Chong, L. & **Yang, M.** 2023 AI vs. Human: The public's perceptions of the design abilities of artificial intelligence. In *Proceedings of the Design Society: International Conference on Engineering Design*, **3**, pp. 495–504.

Cross, N. 2021 *Engineering Design Methods: Strategies for Product Design*. John Wiley & Sons.

Dhariwal, P. & **Nichol, A.** 2021 Diffusion models beat Gans on image synthesis. *Advances in Neural Information Processing Systems* **34**, 8780–8794.

Dinh, L., **Sohl-Dickstein, J.** & **Bengio, S.** 2016 Density estimation using real NVP. arXiv:1605.08803.

Douglas, B. D., **Ewell, P. J.** & **Brauer, M.** 2023 Data quality in online human-subjects research: Comparisons between MTurk, prolific, CloudResearch, qualtrics, and SONA. *PLoS One* **18**(3), e0279720.

Feng, Y., **Wang, X.**, **Wong, K.**, **Wang, S.**, **Lu, Y.**, **Zhu, M.**, **Wang, B.** & **Chen, W.** 2023 PromptMagician: Interactive prompt engineering for text-to-image creation. In *Proceedings of the IEEE Transactions on Visualization and Computer Graphics*, pp. 1–11.

Foster, M. 2021 Design thinking: A creative approach to problem solving. *Management Teaching Review* **6**, 123–140.

Georgiades, A., **Sharma, S.**, **Kipouros, T.** & **Savill, M.** 2019 ADOPT: An augmented set-based design framework with optimisation. *Design Science* **5**, e4.

Giambi, N. & **Lisanti, G.** 2023 Conditioning diffusion models via attributes and semantic masks for face generation. arXiv:2306.00914.

Giannone, G., **Regenwetter, L.**, **Srivastava, A.**, **Gutfreund, D.** & **Ahmed, F.** 2023 Learning from invalid data: On constraint satisfaction in generative models. arXiv:2306.15166.

Golembewski, M. & **Selby, M.** 2010 Ideation decks: A card-based design ideation tool. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, pp. 89–92.

Grathwohl, W., **Chen, R. T.**, **Bettencourt, J.**, **Sutskever, I.** & **Duvenaud, D.** 2018 Ffjord: Free-form continuous dynamics for scalable reversible generative models. arXiv:1810.01367.

Guilford, J. P. 1956 The structure of intellect. *Psychological Bulletin* **53**(4), 267–293.

https://doi.org/10.1017/dsj.2024.51 Published online by Cambridge University Press

Hilliges, O., **Terrenghi, L.**, **Boring, S.**, **Kim, D.**, **Richter, H.** & **Butz, A.** 2007 Designing for collaborative creative problem solving. In *Proceedings of the 6th ACM SIGCHI Conference on Creativity & Cognition*, pp. 137–146.

Ho, J., **Chen, X.**, **Srinivas, A.**, **Duan, Y.** & **Abbeel, P.** 2019 Flow++: Improving flow-based generative models with variational dequantization and architecture design. In *Proceedings of the International Conference on Machine Learning*, pp. 2722–2730.

Hung, W. K. & **Chen, L. L.** 2012 Effects of novelty and its dimensions on aesthetic preference in product design. *International Journal of Design* **6**(2), 81–90.

Jansson, D. G. & **Smith, S. M.** 1991 Design fixation. *Design Studies* **12**(1), 3–11.

Kingma, D. P. & **Dhariwal, P.** 2018 Glow: Generative flow with invertible 1x1 convolutions. *Advances in Neural Information Processing Systems*, 31.

Kudrowitz, B. & **Dippo, C.** 2013 Getting to the novel ideas: exploring the alternative uses test of divergent thinking. In *Proceedings of the International Design Engineering Technical Conferences And Computers And Information In Engineering Conference* 55928, p. V005T06A013.

Kurosu, M. & **Kashimura, K.** 1995 Apparent usability vs. inherent usability: Experimental analysis on the determinants of the apparent usability. In *Proceedings of the Conference Companion on Human Factors in Computing Systems*, 292–293.

Lee, Y. H. & **Chiu, C. Y.** 2023 The impact of AI text-to-image generator on product styling design. In *Proceedings of the International Conference on Human-Computer Interaction*, pp. 502–515.

Li, X., **Su, J.**, **Zhang, Z.** & **Bai, R.** 2021 Product innovation concept generation based on deep learning and Kansei engineering. *Journal of Engineering Design* **32**(10), 559–589.

Linsey, J. S., **Tseng, I.**, **Fu, K.**, **Cagan, J.**, **Wood, K. L.** & **Schunn, C.** 2010 A study of design fixation, its mitigation and perception in engineering design faculty. *ASME Journal of Mechanical Design* **132**(4), 041003.

Lo, C. H., **Ko, Y. C.** & **Hsiao, S. W.** 2015 A study that applies aesthetic theory and genetic algorithms to product form optimization. *Advanced Engineering Informatics* **29**(3), 662–679.

Ma, K., **Grandi, D.**, **McComb, C.** & **Goucher-Lambert, K.** 2023 Conceptual design generation using large language models. In *Proceedings of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* 6, p. V006T06A021.

Mao, X., **Li, Q.**, **Xie, H.**, **Lau, R. Y.**, **Wang, Z.** & **Paul Smolley, S.** 2017 Least squares generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, 2794–2802.

Matejka, J., **Glueck, M.**, **Bradner, E.**, **Hashemi, A.**, **Grossman, T.** & **Fitzmaurice, G.** 2018 Dream lens: Exploration and visualization of large-scale generative design datasets. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12.

Miller, S. R., **Hunter, S. T.**, **Starkey, E.**, **Ramachandran, S.**, **Ahmed, F.** & **Fuge, M.** 2021 How should we measure creativity in engineering design? A comparison between social science and engineering approaches. *ASME Journal of Mechanical Design* **143**(3), 031404.

Miyato, T., **Kataoka, T.**, **Koyama, M.** & **Yoshida, Y.** 2018 Spectral normalization for generative adversarial networks. arXiv:1802.05957.

Mose Biskjaer, M., **Dalsgaard, P.** & **Halskov, K.** 2017 Understanding creativity methods in design. In *Proceedings of the 2017 Conference on Designing Interactive Systems*, pp. 839–851.

https://doi.org/10.1017/dsj.2024.51 Published online by Cambridge University Press

Mukherjee, A. & **Chang, H.** 2023 Managing the creative frontier of generative AI: The novelty-usefulness tradeoff. *California Management Review*.

Nichol, A. Q. & **Dhariwal, P.** 2021 Improved denoising diffusion probabilistic models. In *Proceedings of the International Conference on Machine Learning*, 8162–8171.

Peer, E., **Rothschild, D.**, **Gordon, A.**, **Evernden, Z.** & **Damer, E.** 2022 Data quality of platforms and panels for online behavioral research. *Behavioral Research Methods* **54**, 1643–1662.

Regenwetter, L., **Srivastava, A.**, **Gutfreund, D.** & **Ahmed, F.** 2023 Beyond statistical similarity: Rethinking metrics for deep generative models in engineering design. arXiv: 2302.02913.

Salimans, T., **Goodfellow, I.**, **Zaremba, W.**, **Cheung, V.**, **Radford, A.** & **Chen, X.** 2016 Improved techniques for training GANs. *Advances in Neural Information Processing Systems*, 29.

Schön, D. 1987 *Educating the Reflective Practitioner: Toward a New Design for Teaching and Learning in the Professions*. Jossey-Bass.

Shah, J., **Smith, S.** & **Vargas-Hernandez, N.** 2003 Metrics for measuring ideation effectiveness. *Design Studies* **24**, 111–124.

Short, A. 2023 The Generation of Novel Art Using Collaborative ML Models. In *Proceedings of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference 3B*, V03BT03A065.

Sonderegger, A. & **Sauer, J.** 2010 The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Applied Ergonomics* **41**(3), 403–410.

Tan, B., **Yang, Z.**, **AI-Shedivat, M.**, **Xing, E.** & **Hu, Z.** 2020 Progressive generation of long text with pretrained language models. arXiv:2006.15720.

Wan, Q. & **Lu, Z.** 2023 GANCollage: A GAN-driven digital mood board to facilitate ideation in creativity support. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, 136–146.

Wei, J., **Wang, X.**, **Schuurmans, D.**, **Bosma, M.**, **Ichter, B.**, **Xia, F.**, **Chi, E.**, **Le, Q.** & **Zhou, D.** 2022 Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems* **35**, 24824–24837.

Xie, Y., **Pan, Z.**, **Ma, J.**, **Jie, L.** & **Mei, Q. A.** 2023 Prompt log analysis of text-to-image generation systems. In *Proceedings of the ACM Web Conference*, 3892–3902.

https://doi.org/10.1017/dsj.2024.51 Published online by Cambridge University Press

Appendix

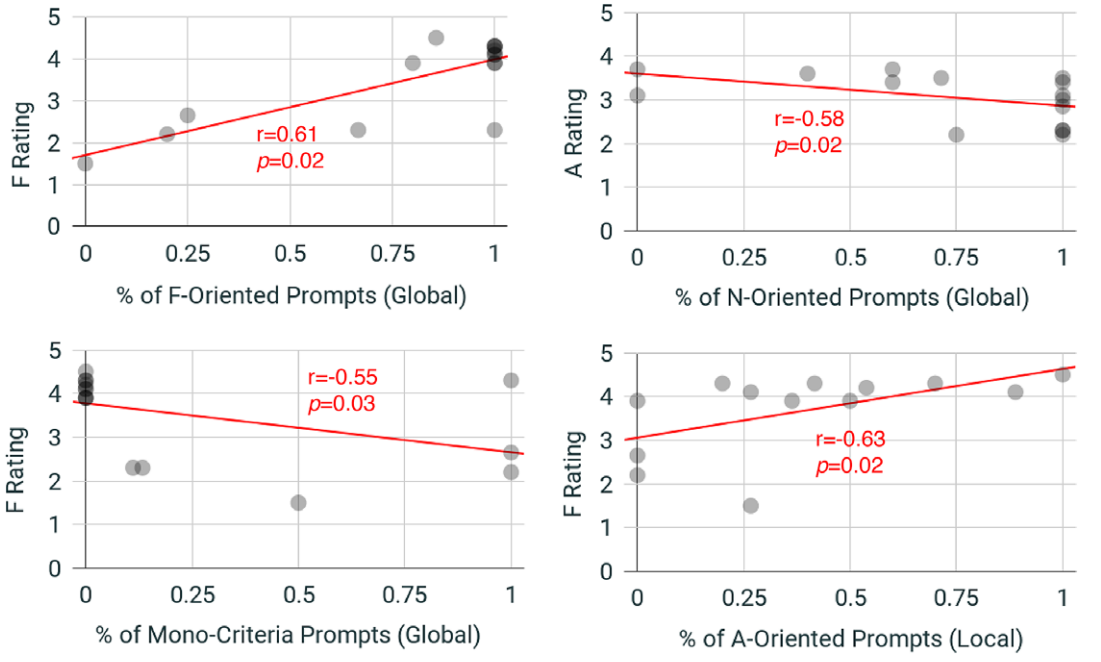

Figure A1. Correlation graphs for all the statistically significant results in Figure 7.

Table A1. Example prompts with different goal orientation(s)

Goal orientation	Example prompt
Feasibility	a bike handle bar with gears
Novelty	a bike with special design
Aesthetics	change the color to gold and keep the current texture
Feasibility & novelty	a wheel with light
Novelty & aesthetics	a bike with crazy design and unique pattern
Feasibility & aesthetics	add chain ring, crank arm and larger pedal
Feasibility & novelty & aesthetics	aesthetically pleasing tiffany-green racing bike with super thick wheels and LED bulb decoration

https://doi.org/10.1017/dsj.2024.51 Published online by Cambridge University Press