Fashioning Creative Expertise with Generative AI: Graphical Interfaces for GAN-Based Design Space Exploration Better Support Ideation Than Text Prompts for Diffusion Models

Richard Lee Davis
École Polytechnique Fédérale de Lausanne (EPFL)
Switzerland
richard.davis@epfl.ch

Thiemo Wambsganss
École Polytechnique Fédérale de Lausanne (EPFL)
Switzerland
theimo.wambsganss@epfl.ch

Wei Jiang
École Polytechnique Fédérale de Lausanne (EPFL)
Switzerland
weijiang9715@gmail.com

Kevin Gonyop Kim
ETH Zürich
Switzerland
kevkim@ethz.ch

Tanja Käser
École Polytechnique Fédérale de Lausanne (EPFL)
Switzerland
tanja.kaeser@epfl.ch

Pierre Dillenbourg
École Polytechnique Fédérale de Lausanne (EPFL)
Switzerland
pierre.dillenbourg@epfl.ch

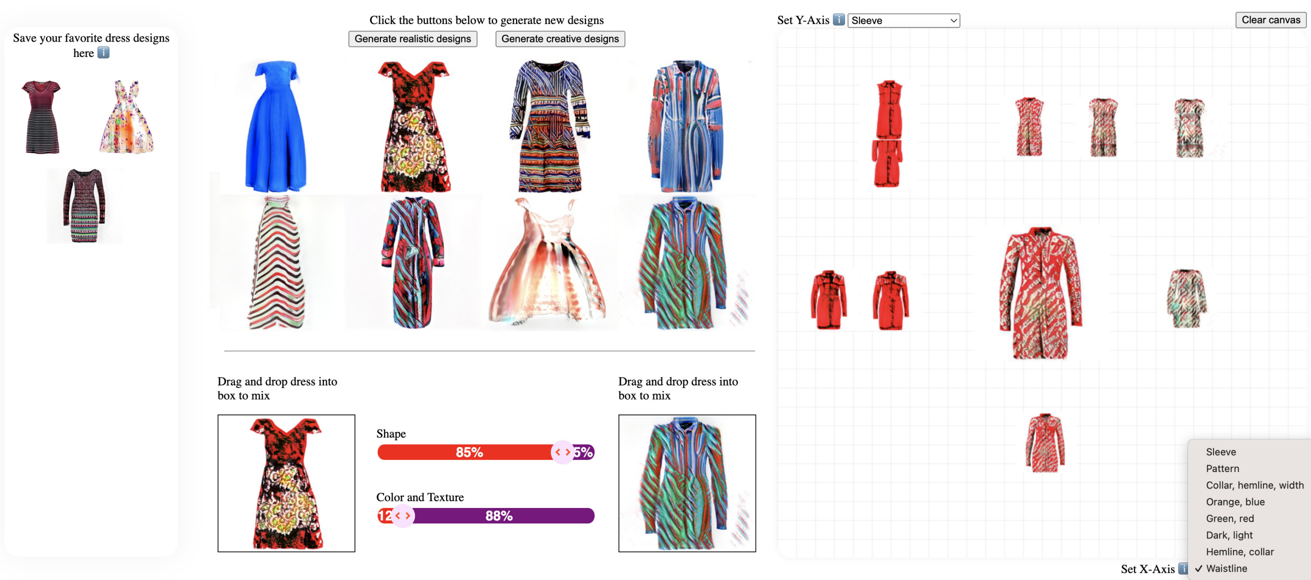

Figure 1: The current version of generative.fashion.

ABSTRACT

This paper investigates the potential impact of deep generative models on the work of creative professionals. We argue that current generative modeling tools lack critical features that would make them useful creativity support tools, and introduce our own tool, generative.fashion[1], which was designed with theoretical principles of design space exploration in mind. Through qualitative studies with fashion design apprentices, we demonstrate how generative.fashion supported both divergent and convergent thinking, and compare it with a state-of-the-art text-based interface using Stable Diffusion. In general, the apprentices preferred generative.fashion, citing the features explicitly designed to support ideation. In two follow-up studies, we provide quantitative results that support and expand on these insights. We conclude that text-only prompts in existing models restrict creative exploration, especially for novices. Our work demonstrates that interfaces which are theoretically aligned with

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642908

[1]A live demo of the tool is available at https://generative.fashion.

principles of design space exploration are essential for unlocking the full creative potential of generative AI.

CCS CONCEPTS

• **Human-centered computing → Interactive systems and tools**; • **Computing methodologies → Artificial intelligence**.

KEYWORDS

Generative AI, Deep Generative Models, Creativity Support Tools (CSTs), Creativity, Design Space Exploration, Fashion Design, Divergent Thinking, Convergent Thinking, Ideation

ACM Reference Format:
Richard Lee Davis, Thiemo Wambsganss, Wei Jiang, Kevin Gonyop Kim, Tanja Käser, and Pierre Dillenbourg. 2024. Fashioning Creative Expertise with Generative AI: Graphical Interfaces for GAN-Based Design Space Exploration Better Support Ideation Than Text Prompts for Diffusion Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA*. ACM, New York, NY, USA, 26 pages. https://doi.org/10.1145/3613904.3642908

1 INTRODUCTION

Deep generative models are neural networks that are capable of creating new things. Recent iterations of these models such as DALL·E [48], GPT-3 [3], and StyleGAN [29] have reached a point where the images, speech, and text that they generate is of such high quality that it is often indistinguishable from original work created by humans. These models typically require an enormous amount of data, compute, and technical know-how to train and run, which has meant that few people outside of academia or industry have been able to access or work with them. However, recently it has become trivial to generate content using large-scale generative models via web portals and APIs (e.g., ChatGPT, DALL·E), and highly-optimized open-source generative models can be downloaded, trained, and run on machines with consumer-grade GPUs (e.g., Stable Diffusion [49], StyleGAN2-ADA [28]).

The wide availability of these models has raised important questions about the nature and future of creative work [8, 47]. While some creative professionals fear job loss and automation, others show enthusiasm towards these new technologies and their potential to provide inspiration, creativity support, or increased productivity [9, 21, 44]. While we are sympathetic with the former group's concerns, we also believe the creative potential of generative AI has immense potential for positive impact. In this paper we argue that unlocking this creative potential requires moving beyond common methods such as random generation and text prompting. More specifically, we hypothesize that alternative modes of interacting with these models may provide better support for ideation.

Ideation is the part of the creative process characterized by both divergent and convergent thinking [12, 13]. Divergent thinking involves the generation of a wide variety of ideas, and convergent thinking involves the selection and refinement of a small set of ideas [15]. Typically, these are viewed as distinct cognitive activities, however both can be characterized as ways of searching through the design space of a domain. The design space of a domain is a "representation of the ideas and concepts that designers develop over time to propose a design solution that materializes into a design artifact" [14, p. 1]. Divergent thinking can be framed as a broad exploration through design space where a large quantity of different ideas are collected, and convergent thinking can be framed as search within a small area of the space that involves the gradual refinement of a small set of ideas.

One of the benefits of framing divergent and convergent thinking in this way is that it helps inform the design of creativity support tools for ideation. In this context, supporting creativity means acting as a guide in the exploration of the design space. To support divergent thinking, the tool should help users identify where they are in the space, allow them to intentionally navigate through the space, and bring them to parts of the space that they are unaware of. These functionalities can help the user break the design fixation which occurs when they become trapped in a small part of the design space [23] and can support them in developing new and surprising ideas [30]. To support convergent thinking, these tools should allow the user to zero in on a specific part of the design space and to explore subtle variations between ideas.

Deep generative models such as GANs [46], VAEs [31], and diffusion models [48] have a unique set of properties that makes them especially well-suited for supporting design space exploration. When trained on large, representative datasets, these models build enormously detailed and complex representations of the design spaces in their learned latent spaces. However, exploration of these latent spaces is difficult due to their high-dimensionality and lack of clear structure. These spaces contain hundreds to thousands of entangled dimensions, which means interpolating an image along a single axis is likely to change multiple properties of the image. Additionally, it can be difficult to intentionally locate a region of space that produces images with desired properties, and even within a very small region of this space there can be an enormous degree of variation.

Common approaches for exploring the latent spaces of generative models involve random sampling or text-prompting interfaces. However, these methods are ill-suited for design space exploration, which necessitates intentional movement through the space rather than random or prompted jumps. These shortcomings can make it difficult for users to holistically explore design options, and can lead designers into restricted corners of the possible design landscape, thereby limiting creativity.

Our tool, generative.fashion, provides technical solutions to these problems and packages them in a web-based graphical user interface that is designed for those with no prior programming experience (Figure 2). The features of generative.fashion were deliberately designed to support styles of design space exploration associated with both divergent and convergent thinking. Multiple ways of locating starting points for exploration are provided, ranging from randomly generating designs to uploading designs and finding the best match in the latent space. Once promising starting points are found, users can continue their exploration of the latent space by mixing designs, which allows users to select multiple points in the space and follow the paths between them, and by using the design canvas, where users can explore the latent space in directions corresponding to sleeve length, pattern, and color by dragging and dropping images in a two-dimensional grid.

We conducted a series of qualitative and quantitative studies to explore how the interaction modalities and underlying features of

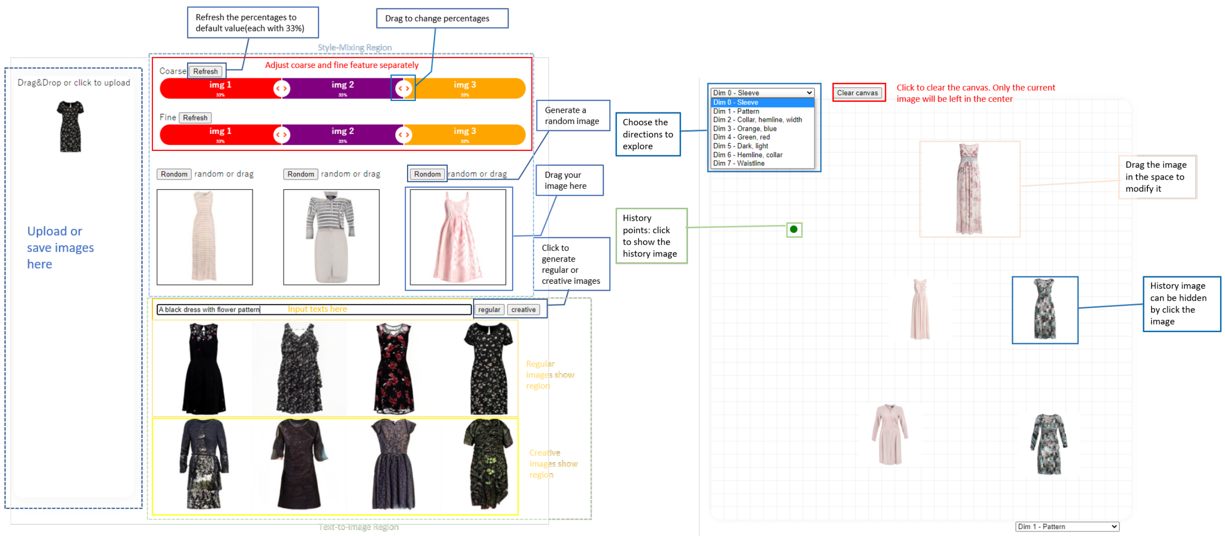

Figure 2: The initial version of the generative.fashion tool. Images could be initially generated via text descriptions using the text box. They could then be dragged to the style-mixing region or saved in the sidebar. Users could selectively combine elements from three designs using the visual style-mixing panel. The output image would be shown in the center of the canvas on the right. The 2D-dimensional canvas represents the design space with two meaningful attributes assigned to the horizontal and vertical axes. These attributes could be changed by using a drop-down menu for each axis. Dragging the image within the canvas was equivalent to moving through the latent space of the GAN in semantically meaningful directions.

generative.fashion might support creativity. In these studies, we compared generative.fashion to both Google Images (a commonly used resource for creative inspiration) and a text-based interface for Stable Diffusion (a state-of-the-art diffusion model) to learn more about how these tools fit into the creative workflow of fashion design apprentices and students from various creative disciplines.

This paper makes the following contributions:

- We introduce generative.fashion, a GAN-based tool designed to explicitly support creative practitioners' ideation via neural design space exploration
- Through a series of qualitative studies, we identify how specific features of generative.fashion support divergent and convergent thinking
- Through a series of quantitative studies, we find that generative.fashion provides significantly better support for creativity than text-based tools
- Based on our findings, we propose a set of principles for the design and use of generative AI for design space exploration

2 RELATED WORK

2.1 Creativity Support Tools for Design Space Exploration

Creativity support tools (CSTs) are broadly defined as digital systems that encompass one or more creativity-focused features which are deployed to positively influence one or more phases of the creative process [12].

The field of CST research has evolved from examining how general-purpose tools might aid creativity [52] to a comprehensive research domain dedicated to developing specialized tools tailored for distinct aspects of the creative process. A notable subset of CSTs includes computer-supported tools for design space exploration. At their best, these tools not only facilitate the discovery of unexplored areas within the design space, fostering divergent thinking, but they also aid in refining and iterating specific designs, supporting the convergent phase of the creative process.

Design space exploration tools can be grouped into four broad categories: parametric exploration, history-based exploration, rule-based exploration, and genetic exploration [51]. Parametric exploration allows generating variations of a design by changing values of parameterized variables [17, 27, 64]; systems with history-based exploration provide a mechanism to keep the history of design changes and to go back in time when needed [32]; rule-based exploration helps the designers explore related examples by suggesting them based on their designs [4, 35]; and genetic exploration involves generating new solutions by combining components of existing designs [30, 50, 56].

However, all of these approaches suffer from the same limitation: they lack a representation of the complete design space. This means that none of these four approaches are capable of supporting all aspects of design space exploration. Parametric exploration can only reach novel regions of the design space through iterative changes to an existing design; history-based approaches can only visit regions that have already been seen; genetic approaches can only explore regions between designs; and example-based approaches can only show points in the design space that correspond to specific examples in their database. Recognizing these limitations highlights the importance of investigating the possibility of neural design space exploration [24], as deep generative models are capable of learning a rich representation of a domain's design space and offer the potential to bring together the best elements from the existing approaches.

2.2 Latent-Space Exploration Tools for Creativity Support

Not all deep generative models provide the same capabilities for latent-space exploration. Some architectures, such as GANs [46] and VAEs [31], are able to construct a disentangled latent space with semantically meaningful directions. By interpolating a design along these semantically meaningful directions, specific characteristics of a design can be changed while holding others constant. Other architectures, such as denoising diffusion models [18], do not learn latent spaces with these properties. To overcome these limitations, diffusion models for image generation are typically trained with both text and images [48], which provides a way to identify points in the latent space that correspond to written descriptions.

A common approach to guiding latent-space exploration of diffusion models is to provide users with better support for text-based search by automating the process of prompt engineering. By utilizing large language models to transform user prompts, these tools are able to produce more diverse [38], surprising [61], and accurate [62] outputs. However, while text prompts provide an excellent way to find points in the latent space, they are less useful for guiding intentional movement through the space, since very small changes to text prompts result in the generation of entirely new images. Methods such as ControlNet [65] and Image-to-Image [42] provide alternative methods for exploring the latent space of diffusion models, such as using sketches, depth maps, uploaded images, and pose data to guide the generation of images. These methods are beginning to be incorporated into creativity support tools [7, 34, 58], and have been shown to provide better support for creativity due to the additional forms of control that they provide over model outputs. However, these methods do not replace text-based search, but work in combination with it, which is due to the underlying architecture of the diffusion models that have been trained with images and text. Because latent-space exploration remains guided, at least in part, by textual descriptions, there is a risk that these tools will fail to adequately support divergent ideation, since it is impossible for a user to describe a region of the design space that they are unaware of.

Because GANs and VAEs contain a semantic latent space, they are less likely to suffer from this limitation. Creativity support tools built on these types of models use features such as random sampling and generation [39, 60], image-guided synthesis [19, 43], blending outputs by interpolating between their points in the latent space [60, 63], and adjusting specific characteristics of a design by moving it along semantically meaningful directions in the latent space [22, 55].

To support both divergent and convergent ideation via design space exploration, all of these features are important. Divergent ideation is supported by random sampling and generation, which provides a means for finding unknown regions of the design space, and by blending images, which makes it possible to explore regions of the design space that lie between points of interest. Convergent ideation is supported by image-guided synthesis, which allows users to find specific points in the space, and by interpolating along semantically meaningful directions, which makes it possible to zero in on specific designs by changing individual characteristics. Additionally, to be useful for creative practitioners, these tools should provide an interface to the underlying model that is easy to use and that doesn't expect technical knowledge. To the best of our knowledge no existing tool combines all of these features with an intuitive interface. Motivated by this, we created generative.fashion, a platform that integrates all these functionalities behind a user-friendly graphical user interface.

3 SYSTEM DESIGN

3.1 The generative.fashion Design Space Exploration Tool

Our primary goal in developing generative.fashion was to create a tool that could support users in intentionally and meaningfully exploring the latent space of a deep generative model. Achieving this goal required us not only to create an interface to the model with novel interaction modalities, but also to develop new features of the model itself. We followed general principles of design as outlined by Shneiderman et al. [53] to ensure that the tool would foster "easy exploration, rapid experimentation, and fortuitous combinations that lead to innovations" (p. 70).

Our starting point was the StyleGAN2-ada model [28] which we trained from scratch on the Feidegger dataset [37]. The most straightforward way to explore the latent space of this model is through "random generation", which involves randomly sampling points in the latent space and using them to generate output images. However, this method is not useful for searching the latent space to find specific designs.

The source code for this model does provide a method for intentionally locating points in the latent space that we call "image search". Also known as "GAN inversion", this method allows a user to input an image and returns a point in the latent space that produces an output similar to the input image. In practice, we found that the GAN inversion function provided in the source code to be slow and the results to be unsatisfying. To correct this, we employed the method described in [1]. Given a starting initialization w, we search for an optimized vector w^* that minimizes the objective loss function that measures the difference between the given image I and the image $G(w)$ generated from w. The optimal w^* is expressed as

$$w^* = \min_w L(w) = \min_w \|f(G(w)) - f(I)\|_2^2 + \lambda_{pix}\|G(w) - I\|_2^2$$

where the loss is calculated as the weighted sum of perceptual loss [26] and pixel-level loss that is pixel-by-pixel MSE loss between two images. This loss noticeably improved our tool's ability to embed out-of-sample examples in the latent space of the GAN, which made it possible for users to upload their own images and use them as starting points for their design space exploration.

We implemented an additional way of locating a point in the GAN latent space by providing a text box where users could write a short description of a design which would be used to locate closely-matching images embedded in the latent space. We call this method "text search". Since this functionality was not a part of the StyleGAN2-ada code, we explored two methods for adding this capability to the model. The first method was to randomly sample images from the latent space, then to pass these along with the text description through a CLIP [45] model to find a small number of images which most closely matched the text. The second method was to fine-tune a DALL·E model [45] on the Feidegger dataset, and then to pass the text descriptions to DALL·E and let it generate designs. Surprisingly, we found that the two methods produced similar results and chose to use the first method because it was more efficient and straightforward.

After users had harvested a crop of images from the GAN latent space using random generation, image search, or text search, they needed a way to begin converging on specific ideas. We provided two distinct functionalities to aid in this process. The first, style mixing, allowed users to blend two images by interpolating between them in latent space. While this functionality was a pre-existing property of the StyleGAN2-ada model [29], we developed a novel user interface to expose the full power of this functionality to the user. Two generated images could be dragged and dropped into the visual style mixing interface, and sliders allowed the user to mix and combine features from each of the designs into a single example. The coarse slider controlled the shape of the output, and the fine slider controlled the pattern and color. The output image was shown in the center of the latent-space exploration panel on the right in Figure 2.

The second was the latent-space exploration panel, which allowed users to take any generated image and move it along meaningful directions in the latent space. This provided users with a way to intentionally and meaningfully explore the latent space of the GAN. Each axis of this two-dimensional canvas corresponded to a semantically-meaningful direction in the latent space, and the direction corresponding to each axis could be changed using a drop-down menu. Dragging and dropping an image within the canvas was equivalent to interpolating a point in the latent space along the directions selected in the drop-down boxes, and each newly-generated point was shown on the canvas as a history point. For a simplified representation of this interface see Figure A1 in the Appendix. Behind the scenes, we used a method described in [16] to identify semantically-meaningful directions using PCA on the latent w space corresponding to sleeve length, pattern, color, hemline, waistline, and more. For more information on this method and for more examples of the results of interpolating along the different principal components, see [25] and Figure A2 in the Appendix.

The GAN was hosted on a Google Cloud Compute virtual machine with 1 NVIDIA T4 GPU, 4 virtual cores, and 15 GB of RAM. This machine was capable of generating 8 images in under a second and could serve requests from over 20 concurrent users without issue.

3.2 The Text-Based Interface for Stable Diffusion

We created an online, text-based interface for Stable Diffusion which accepted a text prompt and generated an image when the Submit button was pressed (Figure 3). For brevity, we refer to this tool as the "Stable Diffusion tool" in the paper. Images were generated using the Stable Diffusion v2-1 base model [49], which was hosted on a Google Cloud Compute virtual machine with 4 NVIDIA L4 GPUs, 48 virtual cores, and 192 GB of RAM. While this machine had roughly 12× the performance of the machine hosting the generative.fashion tool, this was necessary to ensure that it had comparable performance to the generative.fashion tool. The model was built and optimized using the Huggingface diffusers library [59], and was capable of generating an image in 4 seconds (which is near the limit of what is currently achievable) and could serve requests from over 20 concurrent users without issue.

4 EXPERIMENTAL EVALUATION

In order to assess the effectiveness of generative.fashion as a creativity support tool, we followed a mixed-methods approach, where we combined both qualitative insights with quantitative results. This approach allowed us to triangulate our findings and shed more nuanced light on different aspects of deploying generative models in user-centered system for creativity support. Specifically, we conducted two qualitative studies with fashion design apprentices actively engaged in clothing design, followed by two quantitative studies involving a broader sample of students from various creative disciplines. Table 1 provides an overview of our four studies.

Our qualitative studies focused on observing how fashion design apprentices integrated generative.fashion into their existing creative processes, and compared this with Google Image search, the benchmark approach the design apprentices normally use, and the Stable Diffusion tool, a large-scale diffusion model with enormous expressive power. These studies were designed to gain insights into how these tools might affect divergent and convergent thinking in a real-world context. Two quantitative follow-up studies were designed to validate and extend the findings from the qualitative phase. These studies were conducted online, using within-subjects designs, and validated constructs like the Technology Acceptance Model (TAM3) and Creativity Support Index (CSI) were used to measure constructs of interest. Our overall goal in comparing generative.fashion with the Stable Diffusion tool and Google Image was to better understand how the interaction modalities and underlying features of generative.fashion might specifically impact ideation when compared to conventional, text-prompt based methods.

4.1 Qualitative Evaluation with Fashion Design Apprentices

To investigate whether and how generative.fashion might be used to support ideation in an authentic context, we conducted a series of two qualitative studies with a group of seven fashion design apprentices (3F, 4M, ages 17–25) studying in a Western European school for vocational training. The goal of the first study was to investigate the apprentices' current use of technology in their creative practice, to introduce the generative.fashion tool and give them the opportunity to use it in a design problem, to observe how they used the different features of the tool during the design process, and to discuss with them how the tool might fit into their existing creative practice. In the second study, we introduced the apprentices to the Stable Diffusion tool, a text-based diffusion model with enormous expressive power. While generative.fashion lacked the expressiveness of the Stable Diffusion tool, we were curious to see if the features we had built into the tool to support intentional and meaningful design-space exploration might support to the apprentices' ideation process in ways that the diffusion models could

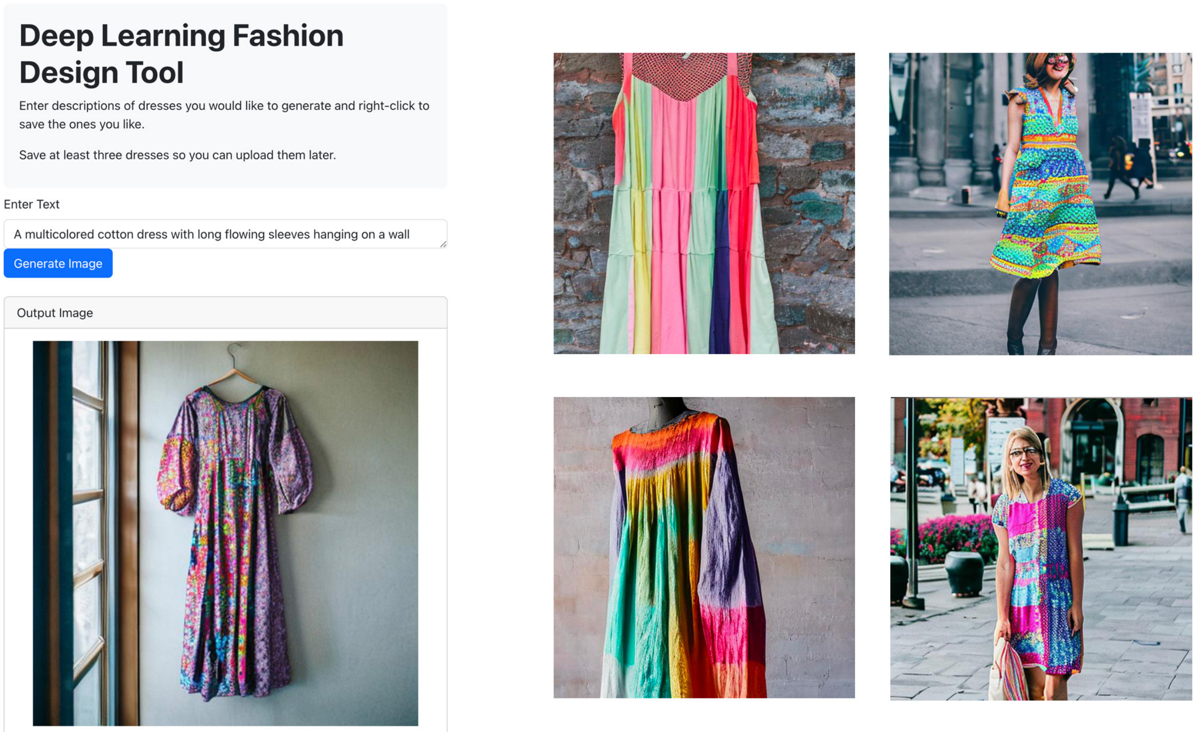

Figure 3: The Stable Diffusion tool interface. On the left, the web interface of the tool is shown, consisting of a text prompt box, a Generate Image button, and an output area for the generated image. On the right, four example images are displayed, each generated using the same text prompt. The two images on the right show common glitches (scrambled faces) produced by the model.

Study	Objective	Design	n	Results
Qualitative Study 1 (Fashion Apprentices)	Investigate how fashion design apprentices integrate generative.fashion into their creative processes	In-person observation and interviews	7	Apprentices favored generative.fashion for divergent thinking
Qualitative Study 2 (Fashion Apprentices)	Compare the effectiveness of generative.fashion with the Stable Diffusion tool in supporting creative practices with fashion design apprentices	In-person observation and interviews	7	Apprentices felt that generative.fashion supported most phases of ideation better
Quantitative Study 1 (Creative Students)	Validate the qualitative findings with a broader population and compare generative.fashion to Google Images	Online experiment	48	Generative.fashion had higher CSI scores and was generally preferred
Quantitative Study 2 (Creative Students)	Extend the comparison to include the Stable Diffusion tool and validate qualitative findings with broader population	Online experiment	39	Consistent preference for generative.fashion over other tools

Table 1: Overview of the Experimental Evaluations in Four Subsequent Studies

not. The details of these two studies are provided in the following sections.

4.1.1 *Qualitative Study 1: User Testing generative.fashion with Fashion Design Apprentices.*

Participants and Procedure. N=7 fashion design apprentices (3F, 4M, ages 17–25) studying in a Western European school for vocational training participated in the study. At the start of the study, the apprentices verbally consented to participate as per the guidelines approved by our university's ethical review board. The apprentices were broken into three groups, with one researcher embedded in each of the groups. The groups were asked to use a collaborative whiteboarding tool to put together a research book containing one or more dress designs for a design persona. Each of the group members took turns using the generative.fashion tool for 10–15 minutes to create dress designs. While one group member was using generative.fashion, the others used their laptops to find and create dress designs using the tools and methods that they would normally use. Finally, after creating the research book, each of the apprentices created a final sketch of a dress for the client in accordance with a set of explicit guidelines[2]. While the apprentices worked on their projects, the researchers took detailed observational field notes on which features of the generative.fashion tool apprentices used and how

[2]The activity and materials were developed in consultation with the fashion design instructor to ensure their authenticity.

their use of these features evolved as the project progressed. Please see Figure 4 for an illustration of the study design and procedure.

Data Collection. During the study each of the researchers led two semi-structured focus groups with the small group of 2–3 apprentices that they were embedded with. The focus group method [33] was used to investigate the apprentices' shared experiences related to technology use in their work as fashion designers, to surface their collective understanding of how the features of the generative.fashion tool might be different from the tools they were already using, and to allow the apprentices to collectively identify salient aspects of the tool. The first focus group took place at the midpoint of the study, after the apprentices had worked with generative.fashion and completed their research books. In this focus group apprentices were asked about the usefulness and usability of the tool, the creativity support provided by the tool, and how the tool compared to other tools and methods. The second focus group took place at the conclusion of the study after they had created their final sketches. In this focus group, the apprentices were asked to explain how their final sketches were influenced by the different tools (including generative.fashion) and asked to describe the activities and settings in which the tool might be most useful. During the focus groups, the researchers took detailed notes and acted as moderators with the goal of ensuring that all apprentices had the opportunity to share their thoughts and perspectives. This moderation approach was intended to compensate for some of the limitations of the focus group method related to individuals dominating the discussion [54]. Shortly after the conclusion of the study, the researchers conducted a debriefing session [41] to share and reflect on emergent findings.

Analysis. The notes from the observation, the focus groups, and the debriefing session were analyzed using a hybrid process of deductive and inductive thematic analysis [10] to identify overarching themes. The notes were first analyzed according to a set of deductive codes derived from our research objectives, and during this deductive coding process emergent themes that surfaced were assigned inductive codes. The lead author was responsible for this coding process, but to ensure that the codes were objective a second author coded 20% of the data using the codebook so that measures of inter-rater reliability could be computed. The raw percent agreement scores were 94%, and Cohen's kappa was 0.63, which according to Fleiss et al. is in the "fair to good" range. Finally, these codes were grouped into a small number of themes that captured important aspects of the participants' experiences and ways of using the generative.fashion tool during the study. We elaborate on these themes in the following sections.

Theme: Better Support for Convergent than Divergent Exploration. Activities associated with the careful refinement of a single idea were coded as convergent, while activities that resulted in the generation of a number of new ideas were coded as divergent. Based on our observations, the apprentices mainly used the tool to support convergent ideation. When using the tool in a convergent manner, apprentices typically started with a single text prompt like "classic gray dress", viewed the resulting image, and then made a series of minor changes to the text in an attempt to tweak the output (e.g., adding the substring "with black buttons"). Then, the apprentices would either make small changes in the style-mixing area or skip the style-mixing area entirely, before moving on to the design canvas to generate a number of designs with minor variations. Each step of the process resulted in further refinements to a dress design, and once the apprentice reached the end of this process they would copy the final dress into the research book.

While all seven of the apprentices primarily used the tool in this manner, two of the apprentices transitioned to using the design canvas feature in a divergent way as the project progressed. Instead of using the design canvas to refine a design, they shifted to dragging the dress across large distances. This resulted in the generation of dresses with strange forms and vivid colors which had little in common with their original designs.

Despite the fact that we did not observe much activity related to divergent ideation, the apprentices viewed the tool as one that could help them find inspiration and produce new ideas. One said, "If you want to mess around and get new ideas and inspiration, …mess around for a few minutes you have something pop up out of nowhere". Another stated that "What I liked about the canvas is depending on where you drag the dress you get things that you haven't really seen in daily life, or in pictures".

Theme: Support for Intentionality and Sense of Ownership. When asked to compare the generative.fashion tool to their existing practices, which were using Internet search engines like Google and Pinterest to look for ideas, some apprentices said that the tool offered support that these existing tools did not. The primary benefit mentioned by apprentices was that generative.fashion made it possible to realize one's own ideas, which provided more of a sense of ownership over the dress designs. One apprentice stated that first-year designers should use this tool instead of Google, since the tool would allow them to "to get their creativity and images in their head more refined". This would allow novices to "build a foundation for their own creativity", allowing them to develop a stronger identity before being influenced by others' work online.

Theme: Proposed changes to the system. The primary criticism voiced by all of the apprentices was that the output image quality was too low. They compared the tool's output to that of Internet search engines, saying "on Google or Pinterest it's possible to find extremely accurate images if you know what to search for". Some apprentices said the images were too small and blurry, while others were unsatisfied with the system's inability to produce fine details such as specific types of buttons or patterns.

4.1.2 Qualitative Study 2: User Testing generative.fashion and the Stable Diffusion tool with Fashion Design Apprentices. Between the first and second study a number of changes were made to the generative.fashion tool to better support divergent ideation and to improve usability. The primary change was that the text-prompt was replaced with two buttons that would randomly generate "traditional" or "creative" designs by sampling from smaller or larger volumes of the latent space. This feature was meant to bootstrap divergent thinking by presenting users with a variety of designs sampled from random points in the GAN latent space at the very start of the design process. Additionally, the style-mixing panel was simplified by reducing the number of designs to be mixed from three to two and the position of different features was rearranged

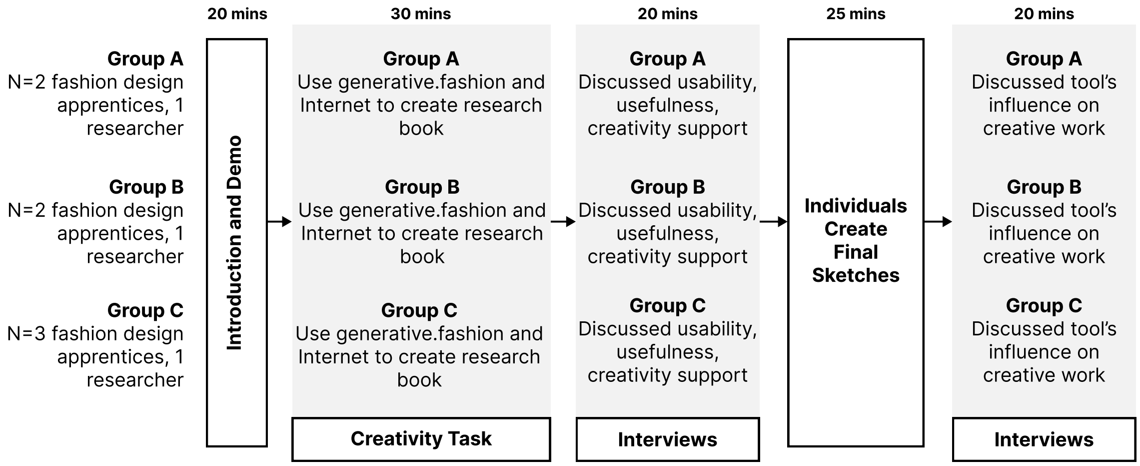

Figure 4: Overview of the design and procedures for the first qualitative study.

to better indicate the intended workflow. The second version of the generative.fashion tool can be seen in Figure 1 and a live demo of the tool can be found at https://generative.fashion.

In addition to evaluating the impact of these changes on the apprentices creative practices, we were also interested in performing a more direct comparison between our tool and the Stable Diffusion tool, an extremely powerful, expressive, and general generative model. Our choice to introduce this tool was motivated by criticisms of generative.fashion related to low image quality and inability to produce highly-accurate images using text prompts. While the Stable Diffusion tool provided solutions to these problems, it lacked features provided by generative.fashion that were specifically designed to support intentional design-space exploration. By comparing the two tools, we hoped to learn whether the unique features of generative.fashion could offer advantages during the ideation process over the more general, text-based interface to the Stable Diffusion model.

Participants and Procedure. The same N=7 fashion design apprentices who took part in the first study also took part in the second study. Like the first study, the apprentices verbally consented to participate as per the guidelines approved by our university's ethical review board. The apprentices were split into two groups, with one researcher embedded within each group. The apprentices within a group did not collaborate with one another, but worked individually on their design collections.

The study took place over three hours. The initial activity that the apprentices took part in was spending 15 minutes sketching an initial design for their collection. After completing this sketch they moved on to working with the tools to generate ideas. One group of apprentices worked with the generative.fashion tool and the other group worked with the diffusion modeling tool. After 45 minutes had passed, the apprentices were asked to stop using the tool and to sketch a new design inspired by their work with the tool.

In the next phase the apprentices swapped tools. If they had been working with the generative.fashion tool, they switched to working with the diffusion model (and vice versa). Again, they spent 45 minutes generating ideas, after which they spent 15 minutes sketching a new design. Finally, all of the apprentices displayed their three drawings on a central table for a gallery walk facilitated by the fashion design instructor. Please see Figure 5 for an illustration of the study design and procedure.

Data Collection. Researchers took observational field notes during the activities, and led three focus group discussions during the study. The first and second focus groups were conducted at the group level and took place after the apprentices worked with one of the tools. These semi-structured discussions were focused on the usefulness and usability of the tool, the creativity support provided by the tool, and how the tool compared to other tools and methods. The final focus group discussion took place with the entire class after the gallery walk. In this discussion apprentices were asked to explain how the different sketches were influenced by the tools, to talk about how the tool might have helped them come up with ideas that they wouldn't have otherwise discovered, and to describe the activity and context in which they would use these tools again.

Analysis. After the study concluded, the researchers debriefed to compare notes and surface insights from their observations and focus group interviews, and the full set of notes were analyzed using the a hybrid process of deductive and inductive thematic analysis. Again, the lead author was responsible for this coding process, but to ensure that the codes were objective a second author coded 20% of the data using the codebook so that measures of inter-rater reliability could be computed. The raw percent agreement scores were 91.75%, and Cohen's kappa was 0.48, which according to Fleiss et al. is in the "fair to good" range. As before, these codes were grouped into a small number of themes which overlapped with the themes from the first qualitative study, but were not identical. We elaborate on these themes in the following sections.

Theme: Supporting Both Divergent and Convergent Exploration. In contrary to the first study where the apprentices mainly used the generative.fashion tool for convergent exploration, in this study we observed a better balance between divergent and convergent exploration approaches. The apprentices used the random-generation

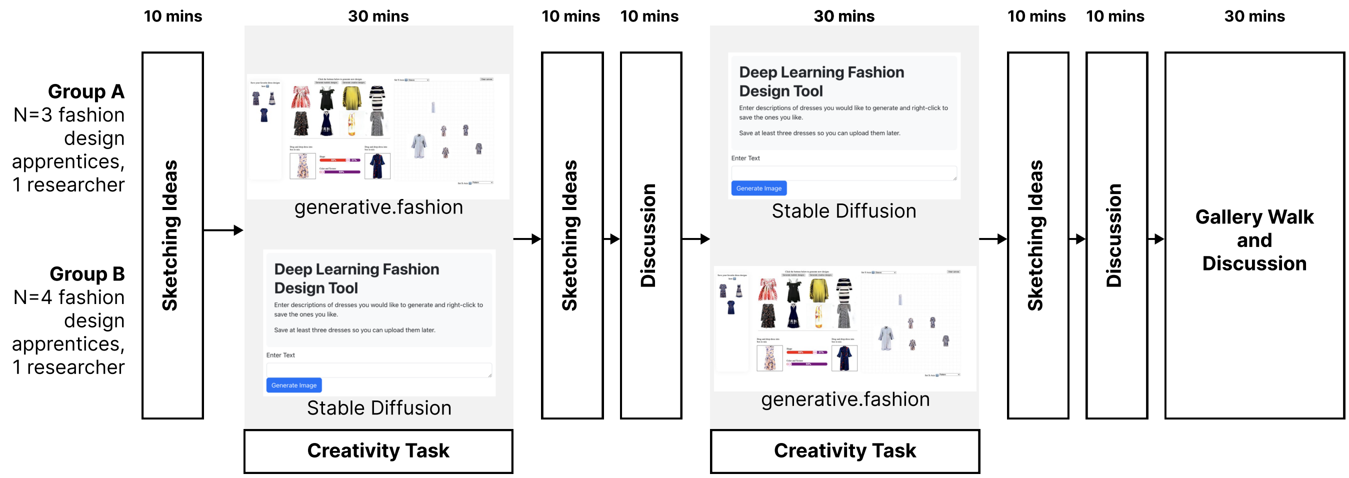

Figure 5: Overview of the design and procedures for the second qualitative study.

functionality to produce wider varieties of dress designs, and then used the design canvas to explore large volumes of the design space (Figure 6). Many of the apprentices stated that the tool's support for divergent exploration was useful in the context of their projects. One apprentice said using the tool "made my brain go places it hadn't gone before", and another said, "it was useful for my project…it helps for getting outside of a thought box". The apprentices agreed that the designs produced by generative.fashion were surprising, saying "it produced mega-creative patterns" and "the tool produced infinitely many possibilities".

Opinions about the tool's support for convergent thinking were more varied. Some apprentices found that the tool offered enough control to hone in on a specific idea, while others were frustrated with their inability to "get where [they] wanted to go".

Theme: Comparing generative.fashion to the Stable Diffusion tool. During the gallery walk, each apprentice took turns presenting their three sketches and explaining which elements were inspired by the use of the different tools. When asked which sketches they were happiest with, six out of the seven apprentices indicated that the drawing created with the generative.fashion tool was their favorite, and stated that they preferred working with generative.fashion over the Stable Diffusion tool. When these apprentices were asked to explain why they preferred using generative.fashion, they provided a number of reasons. Some apprentices valued the ability to explore different patterns and colors without modifying the form of the dress, while others found that the tool made it easier for them to explore different silhouettes and forms. In contrast to generative.fashion, the version of the Stable Diffusion tool used by the apprentices did not provide ways to make these kinds of fine-grained changes to the garments since desired adjustments required the apprentices to input a modified text prompt, which would generate a completely new output image.

A number of apprentices found inspiration in the surprising details generated by generative.fashion, such as pointy shoulders, irregular folds and cuts, and spiky sleeves. Many of these details were included in the apprentices' sketches and were the features that the apprentices liked the most. One apprentice said, "I really like this dress because of the shape and the line down the middle…which was inspired by the tool". In contrast, the apprentices found the outputs produced by the Stable Diffusion tool less inspiring. One apprentice said, "I felt like it was less creative because it was so specific, there wasn't much to change. No room for imagination because it was so accurate". Another said, "when I looked up something I got what I expected, nothing unexpected".

However, one aspect of the Stable Diffusion tool was felt to provide advantages over generative.fashion: the images the Stable Diffusion tool produced were of higher quality and the output was more accurate than generative.fashion. One apprentice stated, "I can be more detailed with specific things like buttons, pockets, colors…It's accurate, you can even look up brands and not recognize any of the pieces but it fit the aesthetic". Another apprentice explained that the Stable Diffusion tool might be useful to further refine specific aspects of ideas created using the generative.fashion tool, saying "If I specifically needed a pocket or sleeve, I would maybe use Stable Diffusion because there are more specific images there".

4.2 Discussion of Qualitative Studies

In the first qualitative study, apprentices worked with a version of generative.fashion which included a text input box that they could use to describe and search for a design in the latent space of the GAN. We primarily observed the apprentices using the tool for convergent ideation, typically following a sequence of actions which began with a text prompt (e.g., "A red dress with long sleeves") and then proceeded through uses of the other features to hone in on the original idea. Only two apprentices were observed who used the tool to generate new ideas in a divergent manner, and these apprentices did so briefly at the very end of the activity. We hypothesized that the text prompt was short-circuiting the divergent ideation process, since this feature forced participants to begin the activity by imagining and describing a specific dress. The primary criticism of the tool was the quality of the outputs, which were described as low-resolution and blurry, and which sometimes bore little resemblance

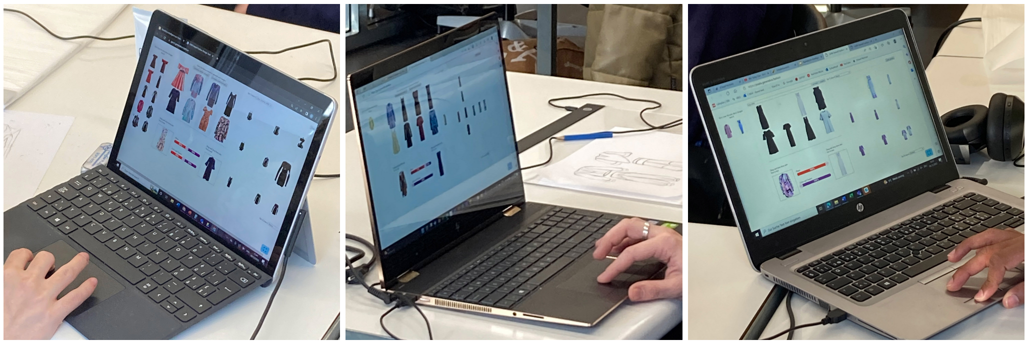

Figure 6: Three different apprentices using the generative.fashion tool to conduct divergent exploration of the design space. The images generated in the design canvas on the right of the screen cover a large area, which corresponds to a large volume of the latent space.

to actual dresses. Despite this, participants described the tool as one that could help them find inspiration and produce new ideas, and that they preferred it over Google Images and other Internet search tools because generative.fashion made it possible to realize their own ideas and provided more of a sense of ownership over the outputs.

In the second study with the same apprentices, we provided them with an updated tool which removed the text prompt entirely. Instead, they were provided with two buttons which would randomly generate dresses by sampling from smaller or larger volumes of the GAN latent space. Additionally, in response to the apprentices' comments about low output quality, we also provided them with access to the Stable Diffusion tool, a more powerful and expressive model which could generate higher-quality images when provided with a text description. We found that removing the text prompt and replacing it with the ability to randomly sample and generate images provided substantially better support for divergent ideation, and observed how apprentices used the tool to smoothly transition from divergent to convergent ideation. In contrast, apprentices found the Stable Diffusion tool to be less useful and less inspiring, and when asked to select their favorite image 6 out of 7 selected images produced by generative.fashion. Apprentices explained that the outputs from the Stable Diffusion tool were less surprising and too accurate. Furthermore, they explained that they lacked the ability to make fine-grained adjustments to specific outputs, since small changes to a text prompt would result in completely new outputs being generated. The only part of the process that the Stable Diffusion tool was preferred for was producing high-quality images of specific details, such as buttons and pockets.

Overall, apprentices expressed a clear preference for generative.fashion over both Google Images and the Stable Diffusion tool. The findings suggested that the reasons for this prefrence was due to the features of generative.fashion, which supported apprentices through the full ideation process, and that tools using text prompts were less useful as they short-circuited the divergent ideation process.

4.3 Quantitative Evaluation with Creative Practitioners

While the qualitative studies offered compelling insights into the design of deep generative creativity support tools, the limited sample size and specificity of the participant pool raised questions about the generalizability of these findings. To increase the robustness and external validity of our initial observations, we next turned to quantitative methods involving a larger and more diverse cohort of creative practitioners.

These follow-up studies were specifically designed to further investigate the following insights that emerged from our qualitative results:

(1) Does generative.fashion offer better support for creative practices than other tools, and in what ways?
(2) Is generative.fashion easier to use and more useful for creative tasks when compared to other tools?
(3) Is generative.fashion more satisfying to use than other tools?
(4) Does generative.fashion provide more control and ownership over outputs than other tools?

To answer these questions, conducted two within-subjects experiments (Figure 7).

The first study compared Google Images to generative.fashion and the second one compared the Stable Diffusion tool to generative.fashion. To answer each of the questions above, we employed the following measures:

(1) The Creativity Support Index (CSI) [5] was used to directly measure support for creativity, and a set of open-response questions were used to investigate support for divergent and convergent ideation. The instrument was unmodified from the original publication.
(2) Two subscales from the TAM3 [57] were used to measure ease-of-use and usefulness (see Appendix A.3)
(3) An instrument based on [6] was used to measure user satisfaction (see Appendix A.3)
(4) A new instrument that we designed, the Design Space Exploration Questionnaire (DSEQ), to try and capture this dimension of the user experience (see Appendix A.2).

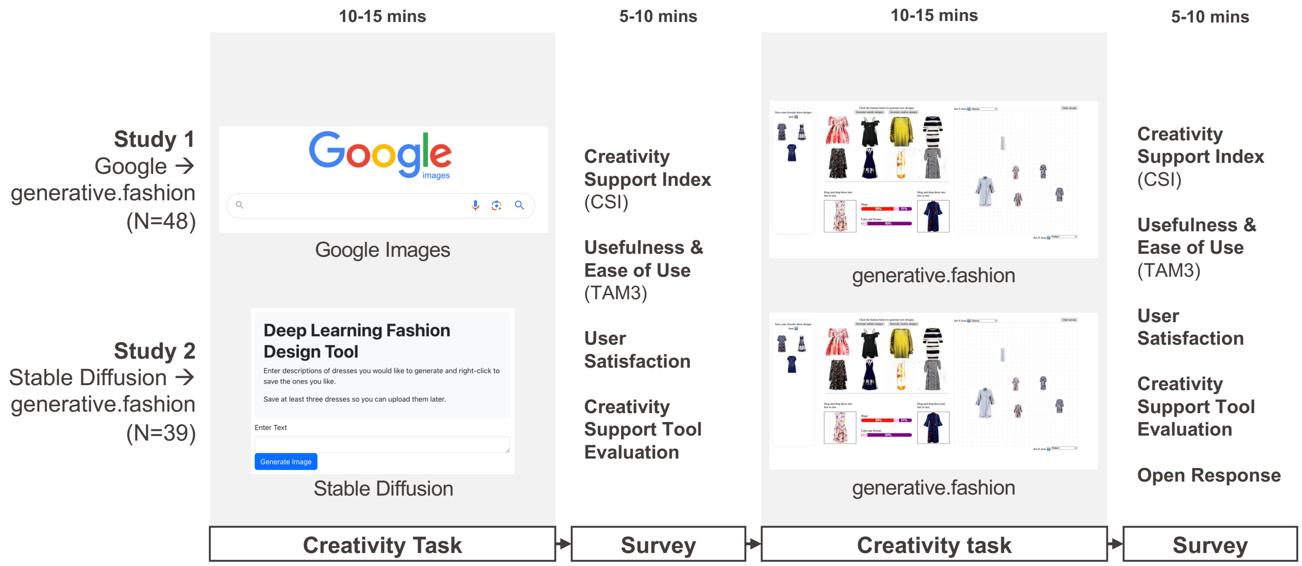

Figure 7: Overview of the experimental design and procedures for the two quantitative studies.

4.3.1 *Quantitative Study 1: Comparing generative.fashion with Google Images.*

Participants. N=48 individuals participated in this online, within-subjects study. Participants were recruited using Prolific (https://www.prolific.co) and were screened based on the following criteria: age between 18 and 35, fluency in English, and either currently studying or having studied one of the following creative domains: Architecture, Art and/or Design, Communication and/or Media, or Fashion and Textiles.

The mean age of the sample was 23.8 years with a standard deviation of 3.20, and the gender distribution consisted of 37 females and 10 males[3]. In terms of racial and ethnic background, the sample was composed of 28 white participants, 8 black participants, 5 of mixed ethnicity, 2 Asian participants, and 4 identifying as "other". Regarding domains of study and practice, 24 were enrolled in Art and/or Design programs, 9 in Architecture, 13 in Communication and/or Media, and 3 in Fashion and Textiles.[4] The sample was geographically diverse, with 30 Europeans (Portugal, Italy, Poland, Germany, Greece, the UK, Spain, and France), 8 from North America (Mexico and the USA), 8 from Africa (South Africa and the Democratic Republic of Congo), and 2 from Australia and 1 from the Philippines. Finally, 38 participants were currently students.

Materials and Study Design. This study employed a within-subjects design to compare the relative effects of two different creativity support tools—Google Images and generative.fashion—on support for creativity, user experience, and design outputs. The rationale for choosing these specific tools was grounded in the findings from the first qualitative study with fashion design apprentices and reflected their actual practices.

Each participant went through two blocks. Each blocks followed a nearly identical procedure, which consisted of an instructional video, a creative task, a questionnaire phase, and open-response questions.

At the beginning of each block, participants read a prompt that framed and introduced the task. This prompt asked them to imagine working for a fashion design agency tasked with creating new dress designs for a client from Zara, and to "come up with a variety of styles with different colors, patterns, and textures…that should be creative, but not too impractical" (the full prompts are available in Appendix A.1). Participants were then introduced to the tool they would be using—first Google Images, then generative.fashion—and then viewed a short instructional video demonstrating how to use the respective tool, save images, and upload designs. They then used the designated tool to search for and save at least three dress designs, which were subsequently uploaded and ranked by the participant.

Following the task, participants completed several questionnaires to evaluate their experience. To directly assess the creativity support of each tool, we employed the Creativity Support Index (CSI) [5]. Additionally, we used specific measures from the Technology Acceptance Model 3 [57] to gauge usability and ease of use, and to assess user satisfaction we administered a measure adapted by Lee and Choi [36] from Chin et al. [6].

Participants also filled out a Design Space Exploration Questionnaire that we created for this study. This questionnaire asked about their satisfaction with the quantity, quality, and diversity of their designs, their pride in their top-three designs, their perceived contribution to the design process (meaningful contribution), their sense

[3]Because one participant declined to provide demographic information, demographic statistics were computed with 47 participants, not 48.
[4]The total count exceeds 47 because some participants were engaged in multiple domains of study.

of control, and the ease with which they realized their ideas (conceptual realization). The full text of each question that participants answered in this instrument can be found in Appendix A.2

Finally, to gather qualitative feedback, participants answered three open-response questions: "What did you like about the tool?", "What about the tool could be improved?", and "Do you have any other feedback, comments, or suggestions about the tool?'

Procedure. Upon entering the study, participants first read an information sheet detailing the study's aims and procedures. Following this, they read and digitally signed a consent form by clicking an opt-in checkbox, as per the guidelines approved by our university's ethical review board. The study then proceeded through the two blocks described in the previous section, with the entire study taking 39 minutes to complete on average. Upon completion, participants received a compensation of 6 GBP for their time and effort.

Data Analysis. For the Creativity Support Index (CSI) questionnaire, we conducted two related analyses. We first calculated the scores on individual dimensions of the CSI instrument using the paired-factor comparisons, and then calculated by taking the weighted average of scores on six distinct dimensions [5]. To investigate differences in overall creativity support between the tools, we conducted a single paired-t-test. Following this, we analyzed each of the constructs in the CSI separately by taking the mean for the items in each of the six constructs (enjoyment, exploration, collaboration, expressiveness, immersion, and results worth effort) and then computing a paired-t-test for each construct.

For the Technology Acceptance Model 3 (TAM3) and user satisfaction measures, we were able to combine the items from each of the three validated constructs (usefulness, ease of use, and satisfaction) into one index per construct by taking the mean. For each tool, this produced one index for usefulness, one index for ease-of-use, and one index for user satisfaction. To look for differences between tools, we conducted paired-t-tests on each of these indices.

For the custom Tool Evaluation questionnaire, which was more exploratory in nature, individual questions were analyzed separately. We conducted seven paired-t-tests to assess specific dimensions such as design quality, diversity, and perceived control. Finally, the qualitative data collected from the open-response questions were analyzed using thematic analysis [2].

The chosen methods for data analysis were designed to offer a comprehensive understanding of the support for creativity and user experience for each tool. By employing validated measures for established constructs, we aimed for robustness in our quantitative assessments. Meanwhile, our exploratory analysis of the DSEQ questionnaire and thematic analysis of qualitative data allowed for a more nuanced interpretation of each tool's relative advantages and disadvantages.

Results: Creativity Support Index (CSI). A paired samples t-test was conducted to compare scores on the Creativity Support Index (CSI) for the two tools, Google Images and generative.fashion. The mean CSI score for Google Images was $M = 65.10, SD = 18.54$, and for generative.fashion, it was $M = 71.9, SD = 17.40$ (Figure 8, left). There was a significant difference in the scores for Google Images and generative.fashion, $t(47) = -2.27, p = .028, d = 0.38$. The overall CSI scores revealed that generative.fashion was rated as significantly more supportive of creativity compared to Google Images.

To further investigate the specific dimensions of creative practice where generative.fashion provided better support compared to Google Images, we conducted paired t-tests for each of the six dimensions. A visualization of these results can be seen in Figure 8, right, and full statistical details can be found in Table 6. When considering the individual dimensions, only the "Expressiveness" dimension showed a significant difference in favor of generative.fashion, while there was a marginally significant difference for "Immersion".

Results: Usefulness, Ease of Use, and User Satisfaction. generative.fashion was rated higher than Google Images in terms of "Usefulness" and "User Satisfaction", with the latter showing a significant difference. Google Images, on the other hand, was rated as significantly easier to use than generative.fashion according to the TAM3 "Ease of Use" measure. See Figure 9 and Table 7 for full details of these analyses.

Results: Design Space Exploration Questionnaire. Participants' experiences with both tools were also assessed using a Design Space Exploration Questionnaire, which consisted of seven questions. Paired samples t-tests were conducted to compare the mean scores for each dimension, and a visual summary of the results can be found in Figure 10. The analysis revealed significant differences between Google Images and generative.fashion in terms of "Quantity", "Meaningful Contribution", and "Control", with generative.fashion outperforming Google Images in these dimensions.

4.3.2 Quantitative Study 2: Comparing generative.fashion with the Stable Diffusion tool. This study was nearly identical to the previous study, with the exception of the tools being compared. To remain concise, we only describe the unique aspects of this study relative to the previous one, and refer the reader to the previous study description for the full details.

Participants. N=39 individuals participated in this online, within-subjects study. Participants were recruited using Prolific (https://www.prolific.co) with the same screening criteria as the previous study.

The mean age of the sample was 25.5 years with a standard deviation of 4.30, and the gender distribution consisted of 27 females and 12 males. In terms of racial and ethnic background, the sample was composed of 29 white participants, 4 black participants, 4 of mixed ethnicity, 1 Asian participant, and 1 identifying as "other". Regarding domains of study and practice, 24 were enrolled in Art and/or Design programs, 7 in Architecture, 11 in Communication and/or Media, and 3 in Fashion and Textiles. The sample was geographically diverse, with 28 Europeans (Portugal, Poland, Italy, the UK, Greece, France, Latvia, the Netherlands, Finland, and the Czech Republic), 3 from the Americas (Argentina, Chile, and the USA), 7 from Africa (South Africa and Zimbabwe), and 1 from Iran. Finally, 28 participants were currently students.

Materials and Study Design. This study employed an identical design to the previous study: a within-subjects design to compare

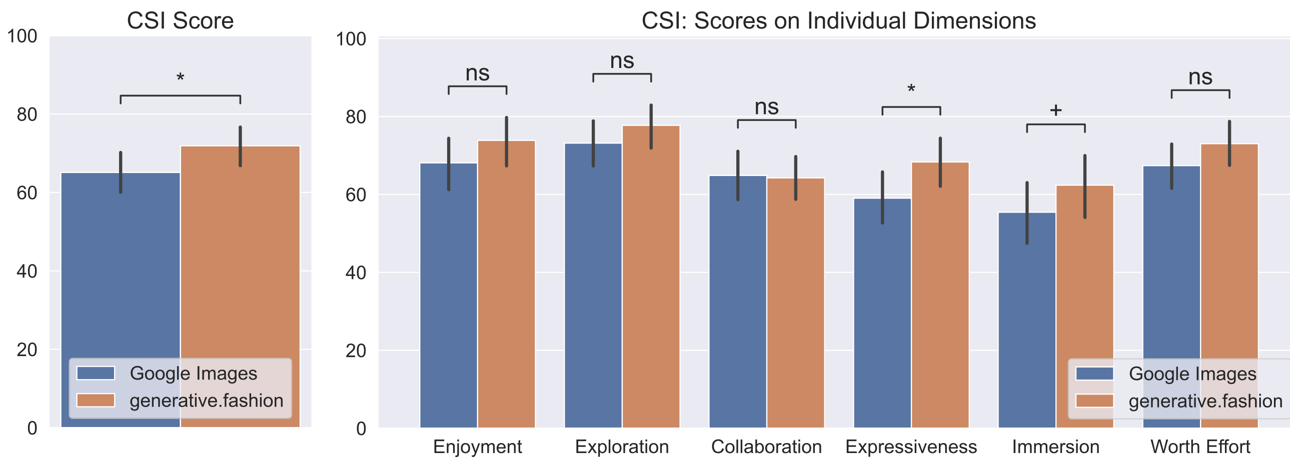

Figure 8: Comparative Analysis of Creativity Support Index (CSI) scores between Google Images and generative.fashion. Left: Barplot of overall CSI scores for the two tools. Right: A set of six barplots illustrating the scores on individual dimensions of the CSI. Error bars indicate standard error, and levels of statistical significance are denoted as follows: $+p < 0.10, *p < 0.05$.

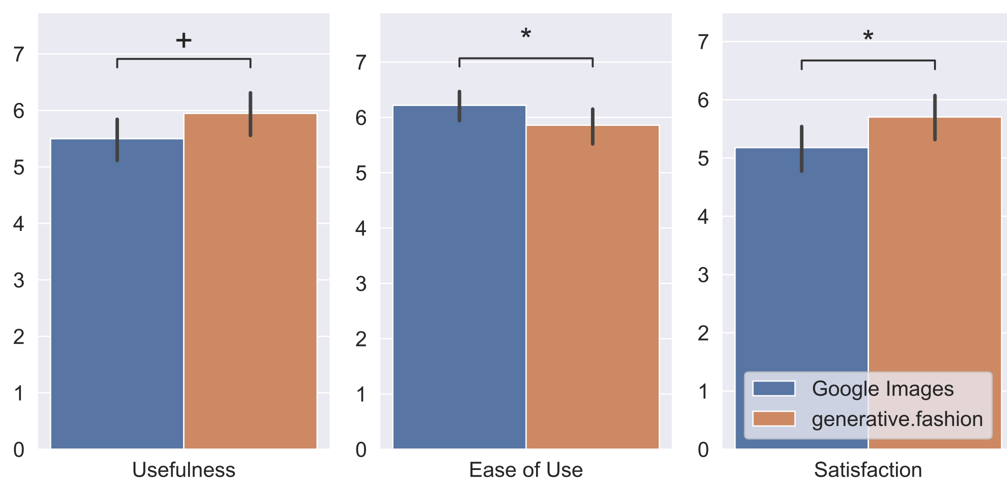

Figure 9: Comparison between Google Images and generative.fashion across three metrics: Usefulness, Ease of Use, and Satisfaction. Each barplot represents scores on a Likert scale from 0 (strongly disagree) to 7 (strongly agree). Error bars indicate standard error, and levels of statistical significance are denoted as follows: $+p < 0.10, *p < 0.05$.

the relative effects of two different creativity support tools—the Stable Diffusion tool and generative.fashion—on support for creativity, user experience, and design outputs (Figure 7). The only difference was that participants used the Stable Diffusion tool, an extremely powerful and expressive generative model capable of producing an enormous variety of detailed images conditioned on text prompts. Otherwise, the participants received the same prompts, were asked to fill out the same set of survey instruments, and were asked the same three open-response questions at the end of the study.

Procedure. The procedure for this study was identical to the procedure in the previous study, which is reported in Section 4.3.1. The entire procedure took 35 minutes to complete on average. Participants received a compensation of 6 GBP for their time and effort.

Data Analysis. The analysis conducted was identical to the previous analysis reported in Section 4.3.1.

Results: Creativity Support Index (CSI). A paired samples t-test was conducted to compare scores on the Creativity Support Index (CSI) for the two tools, the Stable Diffusion tool and generative.fashion. The mean CSI score for the Stable Diffusion tool was $M = 59.0, SD = 23.38$, and for generative.fashion, it was $M = 66.62, SD = 23.21$ (Figure 11, left). There was a significant difference in the scores for the Stable Diffusion tool and generative.fashion, $t(38) = -2.17, p = .036, d = 0.33$. The overall CSI scores revealed that generative.fashion was rated as significantly more supportive of creativity compared to the Stable Diffusion tool.

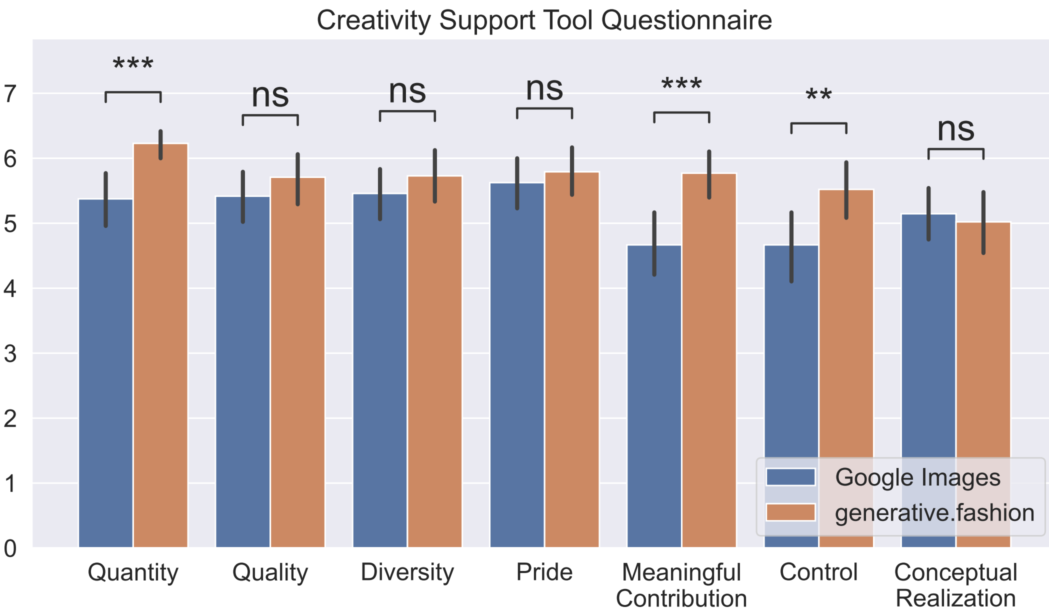

Figure 10: Comparison between Google Images and generative.fashion on the seven questions from the Design Space Exploration Questionnaire. Each barplot represents scores on a Likert scale from 0 (strongly disagree) to 7 (strongly agree). Error bars indicate standard error, and levels of statistical significance are denoted as follows: $+p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001$.

To further investigate the specific dimensions of creative practice where generative.fashion provided better support compared to the Stable Diffusion tool, we conducted paired t-tests for each of the six dimensions. When considering the individual dimensions, both the "Expressiveness" dimension and the "Enjoyment" dimension showed a significant difference in favor of generative.fashion, while there was a marginally significant difference for "Results Worth Effort". A visualization of these results can be seen in Figure 11, right, and full statistical details can be found in Table 9.

Results: Usefulness, Ease of Use, and User Satisfaction. generative.fashion was rated significantly higher than the Stable Diffusion tool in terms of "Usefulness" and "User Satisfaction", while there was no difference on "Ease of Use" (Figure 12). Full statistical details can be found in Table 10.

Results: Design Space Exploration Questionnaire. Participants' experiences with the Stable Diffusion tool and generative.fashion were assessed using the Design Space Exploration Questionnaire. Paired samples t-tests were conducted to compare the mean scores for each dimension. Significant differences were found between the Stable Diffusion tool and generative.fashion in the dimensions of "Quality," "Control," "Pride", and "Conceptual Realization", with generative.fashion outperforming the Stable Diffusion tool in these aspects. Full statistical details can be found in Table 11, and a visual summary of the results can be found in Figure 13.

4.3.3 Results: Open Responses from Quantitative Study 1 and 2. To further investigate our quantitative findings, we analyzed the open responses from all participants. We identified four main themes for users using generative.fashion: Usability and Ease of Use, Creativity Support Capabilities, Quality of Generated Output, and Concerns with Generative AI. The second author was responsible for this coding process, but to ensure that the codes were objective an additional author coded 20% of the data using the codebook so that measures of inter-rater reliability could be computed. The raw percent agreement scores were 95.77%, and Cohen's kappa was 0.55, in the "fair to good" range.

Usability and Ease of Use The first cluster of comments from users dealt with general usability and ease of use of generative.fashion. Participants from both quantitative studies reported that the interface of generative.fashion was generally easy to use. Participants appreciated "the ability to mix different styles together and see the result quickly", "how easy it was to use", and "the easy design of the tool". Additionally, participants reported that "I had to put very little effort", "it was very simple and linear to use", "it was easy and provided a lot of useful options", and "the design made my work faster and more creative". However, some users found specific aspects of the interface to be overwhelming. For instance, one participant noted, "maybe the User interface, it's a bit overwhelming at first glance".

Creativity Support Capabilities We found a second cluster in the responses addressing the creativity support capabilities of generative.fashion. In particular, many comments indicated that the tool's features supported divergent and convergent thinking, in line with the principles of design space exploration. Regarding divergent ideation, students appreciated "the options it allowed me to explore", "the randomness of the designs generated (very fun to play with)", "the amount and variation of options", and "how easy it is

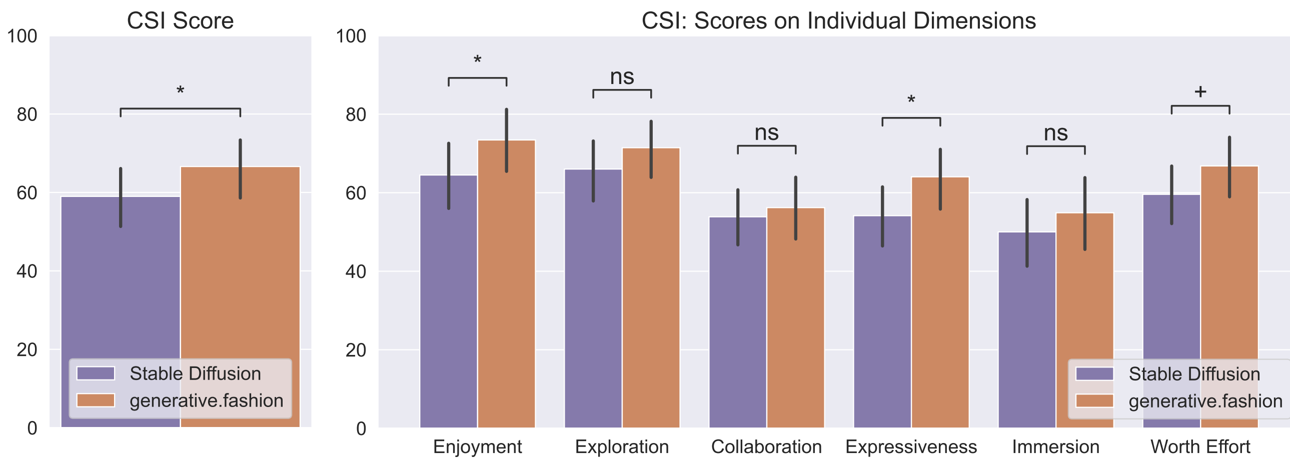

Figure 11: Comparative Analysis of Creativity Support Index (CSI) scores between the Stable Diffusion tool and generative.fashion. Left: Barplot of overall CSI scores for the two tools. Right: A set of six barplots illustrating the scores on individual dimensions of the CSI. Error bars indicate standard error, and levels of statistical significance are denoted as follows: $+p < 0.10, *p < 0.05$.

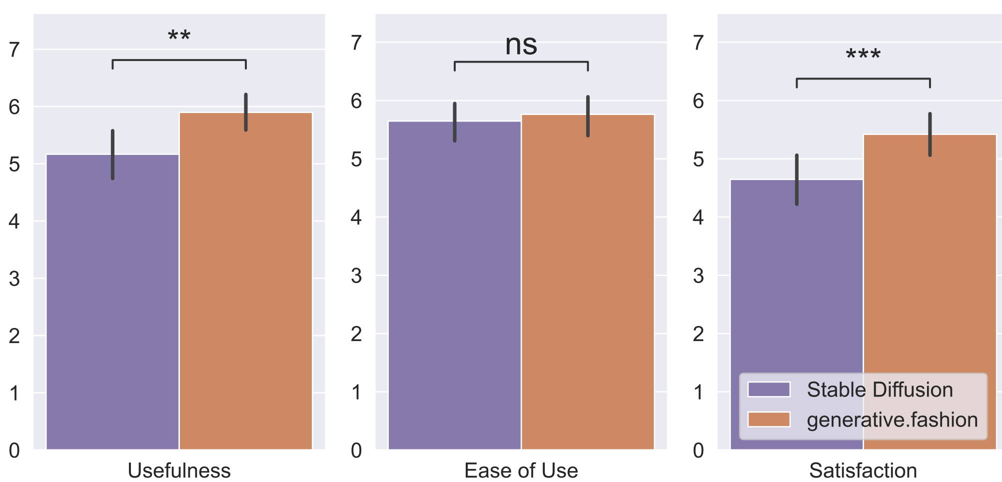

Figure 12: Comparison between the Stable Diffusion tool and generative.fashion across three metrics: Usefulness, Ease of Use, and Satisfaction. Each barplot represents scores on a Likert scale from 0 (strongly disagree) to 7 (strongly agree). Error bars indicate standard error, and levels of statistical significance are denoted as follows: $**p < 0.01, ***p < 0.001$.

to generate random ideas based on an initial design". Additionally, they stated that "it was valuable for creating several ideas spinning from a starting point", that "the random generator was also great to stimulate creativity", and finally that "there were basically endless possibilities". Regarding convergent thinking, participants primarily highlighted the design mixing panel and design canvas as supporting them in honing in on specific designs. Participants stated that the design canvas "gave me the opportunity to change the characteristics of a dress by dragging it to a particular space" and that it was "a creative method of image alteration and served its purpose in successfully editing the images to meet the parameters specified", while the design mixing panel helped participants "customize and explore different dress designs". Many users mentioned that the "stimulated creativity", that the "the design canvas was fun to use" and "gave me more control over how the design came out", and that the design mixing panel provided the ability to "mix and match" designs easily. In general, participants felt the tool enhanced their creativity, saying "I was able to make dresses I wouldn't be able to make normally".

Quality of Generated Outputs The third cluster of comments dealt with the quality of the generated outputs from generative.fashion and the Stable Diffusion tool. Many of the

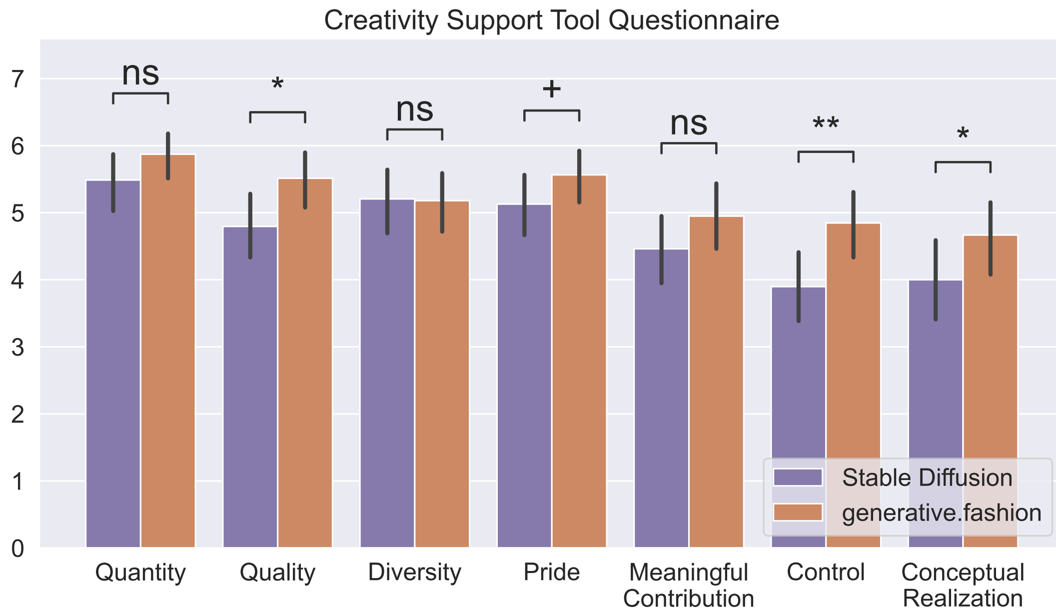

Figure 13: Comparison between the Stable Diffusion tool and generative.fashion on the seven questions from the Design Space Exploration questionnaire. Each barplot represents scores on a Likert scale from 0 (strongly disagree) to 7 (strongly agree). Error bars indicate standard error, and levels of statistical significance are denoted as follows: $+p < 0.10$, $*p < 0.05$, $**p < 0.01$.

participants found these outputs to be strange, off, and too obviously being AI generated. Generally, participants commented that "we can easily see that it's AI" and that "many dresses didn't even look like dresses, they were way too much generated looking". When discussing the Stable Diffusion tool, one wrote that the "deformed faces or extra legs were kinda creepy", and when commenting on generative.fashion participants said "in shape and texture it is still far from the normal", "some of the generated images were out of proportion and some didn't even generate a dress at all", and "sometimes it went too far". However, some participants found these aspects of generative.fashion to be interesting, saying "I liked the abstract aspects of the second tool, the designs were out of the place, but there were something interesting about them". To address these issues, participants provided suggestions such as "there could be a 360 degree view", "it would be easier to find a style of dress if there were filters based on colors, fabrics, lengths, etc." and "Perhaps it could be improved by eliminating designs that are impractical and impractical in reality".

Concerns About Generative AI A small number of participants expressed concerns or made suggestions about how the tool might be used in authentic creative practice. All of these comments expressed the same idea, which was that generative AI should not be designed to replace the designer, but to complement their work. One said "For this task I would still have preferred to sketch something up myself…I would rather use this as part of a process…I do think that working with AI designs further instead of viewing them as an end result will also help make someone feel more included in the creative process". Another expressed this more succinctly, saying "It would be interesting a hybrid intelligence instead of a machine intelligence". Only one expressed anger, saying "A tool should be for refining and detail…This is a try replacement of the artist".

4.4 Discussion of Quantitative Studies

In conducting these experiments we set out to triangulate and validate our qualitative findings, and in both cases we found strong and robust evidence that supported these insights. First, we found that generative.fashion provided better support for creativity than other tools, and in particular that generative.fashion allowed users to be more "creative" and "expressive". Furthermore, our participants found generative.fashion to be more useful and more satisfying to use than other tools by improving performance and productivity and by producing useful and appropriate designs. Finally, participants found generative.fashion to provide more control over how the designs turned out. An analysis of the open-responses corroborated these findings and further supported the qualitative finding that the features of generative.fashion better supported both divergent and convergent ideation than a text prompt interface.

5 GENERAL DISCUSSION

We distill our findings from both the qualitative and quantitative studies into three insights related to the use of deep generative models for creativity support. All of these insights are concerned with different ways of controlling the stochasticity or unexpectedness of the generative model's outputs to support specific types of ideation activity.

5.1 Exposure to unexpected regions of the design space supports divergent ideation

At the start of the divergent thinking phase, we found that the apprentices appreciated when the model produced unpredictable or surprising outputs, and that these sorts of outputs were rarely produced via text prompting. With both generative.fashion and the Stable Diffusion tool, using text prompts appeared to short circuit the divergent ideation process. With the Stable Diffusion tool, the apprentices explicitly said that the outputs produced via text prompting were too accurate, "leaving no room for imagination", and that the model produced "nothing unexpected". And while generative.fashion could not match the accuracy of the Stable Diffusion tool, we observed that apprentices who started by inputting text prompts mostly skipped over the divergent thinking phase entirely. However, after replacing the text-prompt in the generative.fashion tool with buttons for randomly sampling and generating images from the GAN latent space, apprentices engaged in more activities associated with divergent ideation and explicitly stated that the unexpected outputs were inspiring.

These findings indicate that it is important to provide users with ways of rapidly generating multiple outputs from a large volume of the design space, since this will aid them in finding interesting and unexpected regions of the design space in which to continue their exploration. Text prompts may be ill-suited for this task since it is challenging for a user to write a description of a design that they aren't expecting to find. Put differently, it is not reasonable to expect that a user can describe a region of design space that they don't know exists. This insight agrees with recent work which argues that that AI-driven creativity support tools mostly support idea execution in the later stages of the creative process [20].

For generative.fashion, it was trivial to implement features that could expose users to new regions of the design space. This was because the latent space of the underlying generative model closely corresponded to the dress design space, which meant that randomly sampling points from the latent space would reliably produce recognizable dress designs. Implementing such a feature remains an open challenge for high-capacity, general-purpose diffusion models such as Stable Diffusion. While small regions of the Stable Diffusion latent space may correspond to the design space of different domains, it is not clear how to define their location such that randomly sampling from these subspaces would consistently produce images corresponding to a given design space. Paradoxically, this suggests that smaller, less powerful models trained on content from a specific domain may provide better support for divergent ideation than larger, more expressive models.

5.2 Control over model stochasticity supports the transition from divergent to convergent ideation

After identifying promising regions in the design space for further exploration, users should be able to set constraints on model stochasticity that support intentional and meaningful exploration of those regions. For the generative.fashion tool, these constraints took two forms. First, users were able to choose meaningful directions in the design space to explore, and second, they were able to control the size of the steps that they took in these directions. These constraints made it possible for users to intentionally explore individual regions of the design space in the design canvas, and to explore the design space between regions of interest by using the style-mixing panel.

In the design canvas, users were able to select meaningful directions in the design space to explore by assigning properties such as sleeve length, pattern, color, hemline, and neckline to the x- and y-axes. While moving an image along one of these axes, model stochasticity was tightly constrained to only affect the property of the dress that the user wished to change. Additionally, the user could control the amount of stochasticity applied to this property of the dress by moving the image over larger or smaller distances. In practice, we found that the apprentices used these features in two distinct ways. First, apprentices used these features to map out regions of the design space by dragging and dropping images across large areas of the design canvas (see Figure 6 for three examples of how apprentices in Study 2 used the design canvas in this way). Second, they used these features to hone in on a design by exploring small areas in the design canvas, which resulted in increasingly similar designs with small variations.

The version of the Stable Diffusion tool the apprentices used did not provide the ability to intentionally move in meaningful directions of the design space. To change aspects of a design, an apprentice had to tweak the text description and submit this to the model, which would generate an entirely new batch of images. With no way to pin down specific aspects of a generated image such as the form, color, or pattern, intentional exploration of the design space was unpredictable. Apprentices explicitly mentioned this as a downside of the Stable Diffusion tool, and asked for the ability to mix multiple outputs or pin down specific aspects of generated images.

Based on these findings, we argue that it is important to provide ways of exploring the latent space of a generative model by changing specific elements of the model's output without modifying other aspects of the generated image. These features support the transition from divergent to convergent ideation as users move from mapping out a region of the design space to honing in on a specific design.

5.3 High-quality, predictable outputs are more useful for convergent ideation than divergent ideation

Finally, high-quality, precise model outputs are useful during the final phase of the convergent thinking process. In contrast to generative.fashion, the Stable Diffusion tool produced clearer, higher-resolution images, and when provided with detailed prompts it produced designs with more accurate details and styles. Some apprentices stated that this made the Stable Diffusion tool more useful when focusing on smaller details, such as a pocket or a sleeve. Nevertheless, the quality of the output did play as much of a role in the divergent ideation process. In fact, unexpected or glitchy outputs were sometimes seen as inspiring, and the fashion design apprentices incorporated some of these into their final sketches. This suggests that when building generative models to support ideation in visual disciplines, there may be value in trading off fidelity of output for an improved user experience (e.g., speed).

5.4 Limitations and Next Steps

By adopting a mixed-methods approach we were able to gain more confidence in our findings, as the results across all studies were consistent. Nevertheless, there were limitations that should be addressed in future work, and open questions which suggest next steps.

First, in our quantitative studies there were limitations related to the instruments that we used to understand our users' experiences with the different tools. Because none of the validated instruments we used were designed to evaluate generative AI tools, we created our own instrument to measure constructs like satisfaction with output quality and perception of control over the outputs. However, we failed to quantitatively measure other important aspects such as participants' prior experiences, their beliefs, and their feelings about AI. To achieve better coverage of important constructs it will be important to consider the use of newly introduced, exploratory instruments such as the HI-TAM [40] and the MICSI [34] in future work.

Next, while we found substantial qualitative evidence linking different interaction modalities to different types of ideation, we did not provide any quantitative measures directly linking different features to different forms of ideation. One way to address this would be to analyze the interaction data collected by the generative.fashion and Stable Diffusion systems. Using interaction data would provide us with a second source of data related to which features the apprentices used more frequently, the order in which they used different features, and the ways in which they used these features as they moved through the ideation process. Furthermore, by linking their use of these features to their paths through the latent spaces of the different models, it should be possible to identify the features' impacts on divergent and convergent ideation by examining the sizes of steps taken in the latent space.

Another open question has to do with the impact of image quality on creativity support. Comparing generative.fashion to the Stable Diffusion tool indicated that supporting design space exploration is more valuable than producing high quality output, but to definitively answer this question would require a more careful comparison.

Finally, we emphasize that our work is concerned with how different interaction modalities impact convergent and divergent ideation, rather than making claims about which types of deep generative architectures (e.g., GANs vs. diffusion models) are better suited for creativity support. Our work says nothing about the potential impact on creativity of other interaction modalities for diffusion models that are enabled by ControlNet [65], such as sketching and inpainting. If anything, our work highlights this as a promising direction for future research.

6 CONCLUSION

Deep generative models can play an important role in supporting the work of creative professionals, but current tools lack critical features that could unlock their potential. These models are uniquely capable of learning vast and complex representations of design spaces, but users lack intuitive ways of exploring these spaces in intentional and meaningful ways. When augmented with features that provide users with ways of controlling the stochasticity of the model's outputs, these models are better able to support both divergent and convergent ideation. In the generative.fashion tool, we implemented features which constrained model outputs in ways that were aligned with theories of design space exploration and found that these features were successful in supporting creative practitioners throughout the ideation process. Additionally, these features were found to be more important than accuracy and fidelity of output, as evidenced by the clear preferences for generative.fashion over the Stable Diffusion tool for most aspects of the ideation process, despite the fact that the underlying model for generative.fashion was far less powerful. These findings provide support for our hypothesis that unlocking the potential of deep generative models for creative support depends on the development of interfaces and functionalities that are specifically designed to support design space exploration, and provides some assurance that the features we built into generative.fashion were successful in providing this support. We hope to bring attention to our theorized connection between the learned latent space of deep generative models and the design space of a domain, and view the development of tools grounded in this theory as a promising area for future research on creativity support tools and design space exploration.

ACKNOWLEDGMENTS

Thanks to Peter Bühlmann for helping with the logistics of the research, and to the instructors, apprentices, and students who took part in our studies.

REFERENCES

[1] Rameen Abdal, Yipeng Qin, and Peter Wonka. 2019. Image2stylegan: How to Embed Images into the Stylegan Latent Space?. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4432–4441.

[2] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[3] 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.

[4] Kerry Shih-Ping Chang and Brad A Myers. 2012. WebCrystal: understanding and reusing examples in web authoring. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3205–3214.

[5] Erin Cherry and Celine Latulipe. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Transactions on Computer-Human Interaction* 21, 4 (Aug. 2014), 1–25. https://doi.org/10.1145/2617588

[6] John P. Chin, Virginia A. Diehl, and Kent L. Norman. 1988. Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '88)*. Association for Computing Machinery, New York, NY, USA, 213–218. https://doi.org/10.1145/57167.57203

[7] John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: Sketching stories with generative pretrained language models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19.

[8] Ziv Epstein, Aaron Hertzmann, Investigators of Human Creativity, Memo Akten, Hany Farid, Jessica Fjeld, Morgan R Frank, Matthew Groh, Laura Herman, Neil Leach, et al. 2023. Art and the science of generative AI. *Science* 380, 6650 (2023), 1110–1111.

[9] Ramazan Yilmaz Fatma Gizem Karaoglan Yilmaz and Mehmet Ceylan. 2023. Generative Artificial Intelligence Acceptance Scale: A Validity and Reliability Study. *International Journal of Human–Computer Interaction* 0, 0 (2023), 1–13. https://doi.org/10.1080/10447318.2023.2288730

[10] Jennifer Fereday and Eimear Muir-Cochrane. 2006. Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development. *International Journal of Qualitative Methods* 5, 1 (March 2006), 80–92. https://doi.org/10.1177/160940690600500107

[11] Joseph L Fleiss, Bruce Levin, and Myunghee Cho Paik. 2013. *Statistical methods for rates and proportions*. john wiley & sons.

[12] Jonas Frich, Lindsay MacDonald Vermeulen, Christian Remy, Michael Mose Biskjaer, and Peter Dalsgaard. 2019. Mapping the Landscape of Creativity Support Tools in HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–18. https://doi.org/10.1145/3290605.3300619

[13] Jonas Frich, Michael Mose Biskjaer, and Peter Dalsgaard. 2018. Twenty Years of Creativity Research in Human-Computer Interaction: Current State and Future Directions. In *Proceedings of the 2018 Designing Interactive Systems Conference*. 1235–1257.

[14] John Gero and Julie Milovanovic. 2022. Creation and Characterization of Design Spaces. In *DRS2022: Bilbao*. https://doi.org/10.21606/drs.2022.265

[15] Joy Paul Guilford. 1956. The structure of intellect. *Psychological bulletin* 53, 4 (1956), 267. https://psycnet.apa.org/journals/bul/53/4/267/ Publisher: American Psychological Association.

[16] Erik Härkönen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. 2020. Ganspace: Discovering Interpretable Gan Controls. *Advances in Neural Information Processing Systems* 33 (2020), 9841–9850.

[17] Björn Hartmann, Loren Yu, Abel Allison, Yeonsoo Yang, and Scott R Klemmer. 2008. Design as exploration: creating interface alternatives through parallel authoring and runtime tuning. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*. 91–100.

[18] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems* 33 (2020), 6840–6851.

[19] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. 2018. Multimodal unsupervised image-to-image translation. In *Proceedings of the European conference on computer vision (ECCV)*. 172–189.

[20] Angel Hsing-Chi Hwang. 2022. Too late to be creative? AI-empowered tools in creative processes. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–9.

[21] Nanna Inie, Jeanette Falk, and Steve Tanimoto. 2023. Designing Participatory AI: Creative Professionals' Worries and Expectations about Generative AI. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–8. https://doi.org/10.1145/3544549.3585657

[22] Tansin Jahan, Yanran Guan, and Oliver Van Kaick. 2021. Semantics-Guided Latent Space Exploration for Shape Generation. In *Computer Graphics Forum*, Vol. 40. Wiley Online Library, 115–126.

[23] 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

[24] Wei Jiang, Richard Lee Davis, Kevin Gonyop Kim, and Pierre Dillenbourg. 2021. Neural Design Space Exploration. In *35th Conference on Neural Information Processing Systems (NeurIPS 2021)*.

[25] Wei Jiang, Richard Lee Davis, Kevin Gonyop Kim, and Pierre Dillenbourg. 2022. GANs for All: Supporting Fun and Intuitive Exploration of GAN Latent Spaces. In *NeurIPS 2021 Competitions and Demonstrations Track*. PMLR, 292–296.

[26] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In *European Conference on Computer Vision*. Springer, 694–711.

[27] John R. Josephson, Balakrishnan Chandrasekaran, Mark Carroll, Naresh Iyer, Bryon Wasacz, Giorgio Rizzoni, Qingyuam Li, and David A. Erb. 1998. An Architecture for Exploring Large Design Spaces. *AAAI/IAAI* 143150 (1998).

[28] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. 2020. Training Generative Adversarial Networks with Limited Data. *Advances in Neural Information Processing Systems* 33 (2020), 12104–12114.

[29] Tero Karras, Samuli Laine, and Timo Aila. 2019. A Style-Based Generator Architecture for Generative Adversarial Networks. *arXiv:1812.04948 [cs, stat]* (March 2019). arXiv:1812.04948 [cs, stat]

[30] Kevin Gonyop Kim, Richard Lee Davis, Alessia Eletta Coppi, Alberto Cattaneo, and Pierre Dillenbourg. 2022. Mixplorer: Scaffolding Design Space Exploration through Genetic Recombination of Multiple Peoples' Designs to Support Novices' Creativity. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3491102.3501854

[31] Diederik P. Kingma and Max Welling. 2013. Auto-Encoding Variational Bayes. *arXiv preprint arXiv:1312.6114* (2013). arXiv:1312.6114

[32] Scott R Klemmer, Michael Thomsen, Ethan Phelps-Goodman, Robert Lee, and James A Landay. 2002. Where do web sites come from? Capturing and interacting with design history. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1–8.

[33] Richard A. Krueger. 2014. *Focus Groups: A Practical Guide for Applied Research*. Sage publications.

[34] Tomas Lawton, Francisco J Ibarrola, Dan Ventura, and Kazjon Grace. 2023. Drawing with Reframer: Emergence and Control in Co-Creative AI. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. 264–277.

[35] Brian Lee, Savil Srivastava, Ranjitha Kumar, Ronen Brafman, and Scott R Klemmer. 2010. Designing with interactive example galleries. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2257–2266.

[36] SeoYoung Lee and Junho Choi. 2017. Enhancing User Experience with Conversational Agent for Movie Recommendation: Effects of Self-Disclosure and Reciprocity. *International Journal of Human-Computer Studies* 103 (July 2017), 95–105. https://doi.org/10.1016/j.ijhcs.2017.02.005

[37] Leonidas Lefakis, Alan Akbik, and Roland Vollgraf. 2018. Feidegger: A Multi-Modal Corpus of Fashion Images and Descriptions in German. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

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

[39] Ninon Lizé Masclef and Adrien Chuttarsing. 2023. Latent Organism: Embodied Co-Creation with AI. In *Proceedings of the 15th Conference on Creativity and Cognition*. 239–242.

[40] Yaoli Mao, Janet Rafner, Yi Wang, and Jacob Sherson. 2023. A Hybrid Intelligence Approach to Training Generative Design Assistants: Partnership Between Human Experts and AI Enhanced Co-Creative Tools. In *HHAI 2023: Augmenting Human Intellect*. IOS Press, 108–123.

[41] Shannon A. McMahon and Peter J. Winch. 2018. Systematic Debriefing after Qualitative Encounters: An Essential Analysis Step in Applied Qualitative Research. *BMJ Global Health* 3, 5 (Sept. 2018), e000837. https://doi.org/10.1136/bmjgh-2018-000837

[42] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. 2021. Sdedit: Guided image synthesis and editing with stochastic differential equations. *arXiv preprint arXiv:2108.01073* (2021).

[43] Mohammad Amin Mozaffari, Xinyuan Zhang, Jinghui Cheng, and Jin LC Guo. 2022. GANSpiration: Balancing Targeted and Serendipitous Inspiration in User Interface Design with Style-Based Generative Adversarial Network. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–15.

[44] Njeri Ngaruiya, Jonathan Donner, Joshua Kinuthia Baru, and Babra Wanjiku Chege. 2023. The domestication of AI by Kenyan digital creators. In *Proceedings of the 4th African Human Computer Interaction Conference*. 71–75.

[45] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, and Jack Clark. 2021. Learning Transferable Visual Models from Natural Language Supervision. In *International Conference on Machine Learning*. PMLR, 8748–8763.

[46] Alec Radford, Luke Metz, and Soumith Chintala. 2015. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv preprint arXiv:1511.06434* (2015). arXiv:1511.06434

[47] Janet Rafner, Roger E Beaty, James C Kaufman, Todd Lubart, and Jacob Sherson. 2023. Creativity in the age of generative AI. *Nature Human Behaviour* 7, 11 (2023), 1836–1838.

[48] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. Zero-Shot Text-to-Image Generation. In *International Conference on Machine Learning*. PMLR, 8821–8831.

[49] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis With Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10684–10695.

[50] Jimmy Secretan, Nicholas Beato, David B D'Ambrosio, Adelein Rodriguez, Adam Campbell, Jeremiah T Folsom-Kovarik, and Kenneth O Stanley. 2011. Picbreeder: A case study in collaborative evolutionary exploration of design space. *Evolutionary computation* 19, 3 (2011), 373–403.

[51] Naghmi Shireen. 2020. Bridging the gap between design space exploration and generative design interfaces: An exploratory study. (2020).

[52] Ben Shneiderman. 2001. Supporting Creativity with Advanced Information-Abundant User Interfaces. In *Frontiers of Human-Centered Computing, Online Communities and Virtual Environments*. Springer, 469–480.

[53] Ben Shneiderman, Gerhard Fischer, Mary Czerwinski, Mitch Resnick, Brad Myers, Linda Candy, Ernest Edmonds, Mike Eisenberg, Elisa Giaccardi, Tom Hewett, Pamela Jennings, Bill Kules, Kumiyo Nakakoji, Jay Nunamaker, Randy Pausch, Ted Selker, Elisabeth Sylvan, and Michael Terry. 2006. Creativity Support Tools: Report From a U.S. National Science Foundation Sponsored Workshop. *International Journal of Human–Computer Interaction* 20, 2 (May 2006), 61–77. https://doi.org/10.1207/s15327590ijhc2002_1

[54] Janet Smithson. 2000. Using and Analysing Focus Groups: Limitations and Possibilities. *International Journal of Social Research Methodology* 3, 2 (Jan. 2000), 103–119. https://doi.org/10.1080/136455700405172

[55] Koray Tahiroğlu, Miranda Kastemaa, and Oskar Koli. 2021. Ganspacesynth: A hybrid generative adversarial network architecture for organising the latent space using a dimensionality reduction for real-time audio synthesis. In *Conference on AI Music Creativity*.

[56] Michela Turrin, Peter von Buelow, and Rudi Stouffs. 2011. Design Explorations of Performance Driven Geometry in Architectural Design Using Parametric

Modeling and Genetic Algorithms. *Advanced Engineering Informatics* 25, 4 (Oct. 2011), 656–675. https://doi.org/10.1016/j.aei.2011.07.009

[57] Viswanath Venkatesh and Hillol Bala. 2008. Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences* 39, 2 (May 2008), 273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

[58] Mathias Peter Verheijden and Mathias Funk. 2023. Collaborative Diffusion: Boosting Designerly Co-Creation with Generative AI. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–8.

[59] Patrick von Platen, Suraj Patil, Anton Lozhkov, Pedro Cuenca, Nathan Lambert, Kashif Rasul, Mishig Davaadorj, and Thomas Wolf. 2022. Diffusers: State-of-the-art diffusion models. https://github.com/huggingface/diffusers.

[60] Qian Wan and Zhicong Lu. 2023. Investigating Semantically-enhanced Exploration of GAN Latent Space via a Digital Mood Board. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–5.

[61] Sitong Wang, Savvas Petridis, Taeahn Kwon, Xiaojuan Ma, and Lydia B Chilton. 2023. PopBlends: Strategies for conceptual blending with large language models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[62] Yunlong Wang, Shuyuan Shen, and Brian Y Lim. 2023. RePrompt: Automatic Prompt Editing to Refine AI-Generative Art Towards Precise Expressions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–29.

[63] Di Wu, Zhiwang Yu, Nan Ma, Jianan Jiang, Yuetian Wang, Guixiang Zhou, Hanhui Deng, and Yi Li. 2023. StyleMe: Towards Intelligent Fashion Generation with Designer Style. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.

[64] Loutfouz Zaman, Wolfgang Stuerzlinger, Christian Neugebauer, Rob Woodbury, Maher Elkhaldi, Naghmi Shireen, and Michael Terry. 2015. Gem-ni: A system for creating and managing alternatives in generative design. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 1201–1210.

[65] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding Conditional Control to Text-to-Image Diffusion Models.

A APPENDICES

A.1 Activity Prompts for Quantitative Studies

A.1.1 Activity Prompt for the First Design Activity. Imagine you are working for a fashion design agency. One of your clients is a buyer for Zara, an apparel retailer with multiple stores around the world. One day, the client walks in unannounced and asks to see your company's latest dress designs. Your boss shows them a book of designs from last year, but the client is unimpressed, asking, "I've seen all of these, don't you have something new for me?"

Your boss quickly pulls you aside, whispering, "I can stall the client for about 15 minutes. I know it's crazy, but can you quickly come up with some new dress designs? Try to come up with a variety of styles with different colors, patterns, and textures. They should be creative, but not too impractical. The more options she has to choose from, the better, because if this client picks even one of our designs it will be a huge win for our agency!"

As soon as your boss leaves with the client, you turn to you coworker and ask, "This is impossible! How am I supposed to do this?" Your coworker says, "Don't worry. I just found out about a tool that makes it possible to come up with new dress designs quickly. I'll show you how to use it."

A.1.2 Activity prompt for the second design activity. When your boss comes back and takes a look at the designs, she notices one that she's seen before. "Oh no!" your boss says. "We can't use any of the dress designs you created because they might be protected by copyright laws!"

Your boss tells you "I can stall the client for 15 more minutes. You'll have to use an experimental tool that we've been developing to generate new dress designs. This tool uses artificial intelligence so it doesn't always produce perfect designs, but I'm sure you can use it to create dresses with a variety of styles with different colors, patterns, and textures. Remember, they should be creative, but not too impractical. The more options she has to choose from, the better!"

A.2 The Design Space Exploration Questionnaire

	Strongly disagree	Disagree	Slightly disagree	Neither agree or disagree	Slightly agree	Agree	Strongly agree
I am satisfied with the number of designs I generated							
I am satisfied with the overall quality of the designs I generated							
I am satisfied with the overall quality of the designs I generated							
I am proud of my top three designs							
I played an important role in how the designs turned out							
I had a lot of control over how the designs turned out							
When I had an idea for a dress, I was able to easily produce a design which captured my idea							

Table 2: The Design Space Exploration Questionnaire designed for this study.

A.3 TAM3 and User Satisfaction Subscales

	Strongly disagree	Disagree	Slightly disagree	Neither agree or disagree	Slightly agree	Agree	Strongly agree
Using the system improved my performance on the task							
Using the system on the task increased my productivity							
Using the system enhanced my effectiveness on the task							
I found the system to be useful on this task							

Table 3: TAM3 subscale for usefulness from [57]

	Strongly disagree	Disagree	Slightly disagree	Neither agree or disagree	Slightly agree	Agree	Strongly agree
My interaction with the system was clear and understandable							
Interacting with the system did not require a lot of my mental effort							
I found the system to be easy to use							
I found it easy to get the system to do what I wanted it to do							

Table 4: TAM3 subscale for ease of use from [57]

	Strongly disagree	Disagree	Slightly disagree	Neither agree or disagree	Slightly agree	Agree	Strongly agree
Interacting with the system gave me useful designs							
I am satisfied with using the system because it is easier than creating designs myself							
I feel that using the system made me more like an expert							
The designs produced using the system were appropriate for the task							
My overall experience of using the system was satisfactory							

Table 5: User satisfaction survey adopted from [36]

A.4 Statistical Results from Quantitative Study 1: Google Image Search vs. generative.fashion

	Google Images		generative.fashion		t-test		
	M	**SD**	**M**	**SD**	**t(49)**	**p**	**d**
Enjoyment	68.10	28.83	73.88	21.31	-1.49	0.14	0.26
Exploration	73.17	21.09	77.73	18.48	-1.36	0.18	0.23
Collaboration	64.88	23.60	64.25	19.97	0.20	0.84	0.03
Expressiveness	59.04	22.77	68.33	21.97	-2.45	**0.02***	0.42
Immersion	55.44	28.07	62.38	27.52	-1.96	**0.56+**	0.25
Results Worth Effort	67.40	20.74	73.02	19.73	-1.57	0.12	0.28

Table 6: Statistics for the comparative analysis of Google Images with generative.fashion for the individual subscales of the CSI. Levels of statistical significance are denoted as follows: $+p < 0.10$, $*p < 0.05$.

	Google Images		generative.fashion		t-test		
	M	**SD**	**M**	**SD**	**t(47)**	**p**	**d**
Usefulness	5.50	1.34	5.95	1.33	-1.77	**0.08+**	0.34
Ease of Use	6.22	0.95	5.86	1.13	2.07	**0.04***	0.34
Satisfaction	5.18	1.35	5.70	1.34	-2.29	**0.03***	0.39

Table 7: Statistics for the comparative analysis of Google Images with generative.fashion on usefulness, ease of use, and satisfaction. Levels of statistical significance are denoted as follows: $+p < 0.10$, $*p < 0.05$.

	Google Images		generative.fashion		t-test		
	M	**SD**	**M**	**SD**	**t(47)**	**p**	**d**
Quantity	5.38	1.50	6.23	0.78	-4.06	**<0.001*****	0.716
Quality	5.42	1.40	5.71	1.38	-1.33	0.189	0.210
Diversity	5.46	1.38	5.73	1.44	-1.12	0.268	0.192
Pride	5.62	1.39	5.79	1.32	-0.75	0.455	0.123
Meaningful Contribution	4.67	1.83	5.77	1.26	-4.02	**<0.001*****	0.704
Control	4.67	1.87	5.52	1.56	-2.97	**0.005****	0.496
Conceptual Realization	5.15	1.44	5.02	1.64	0.47	0.640	0.081

Table 8: Statistics for the comparative analysis of Google Images with generative.fashion on the DSEQ questions. Levels of statistical significance are denoted as follows: $**p < 0.01$, $***p < 0.001$.

A.5 Statistical Results from Quantitative Study 2: Text-Based Stable Diffusion Interface vs. generative.fashion

	Stable Diffusion		**generative.fashion**		**t-test**		
	M	**SD**	**M**	**SD**	**t(38)**	**p**	**d**
Enjoyment	64.51	26.37	73.44	25.71	-2.08	**0.045***	0.34
Exploration	66.03	25.58	71.46	24.12	-1.52	0.14	0.22
Collaboration	53.87	23.16	56.21	25.59	-0.67	0.51	0.10
Expressiveness	54.15	25.53	64.05	25.21	-2.44	**0.019***	0.39
Immersion	50.03	28.53	54.90	30.44	-1.18	0.25	0.17
Results Worth Effort	59.59	25.35	66.82	24.99	-1.87	**0.067+**	0.29

Table 9: Statistics for the comparative analysis of the Stable Diffusion tool with generative.fashion for the individual subscales of the CSI. Levels of statistical significance are denoted as follows: $+p < 0.10, *p < 0.05$.

	Stable Diffusion		**generative.fashion**		**t-test**		
	M	**SD**	**M**	**SD**	**t(38)**	**p**	**d**
Usefulness	5.17	1.38	5.90	1.01	-3.33	**0.002****	0.60
Ease of Use	5.65	1.04	5.76	1.09	-0.56	0.58	0.11
Satisfaction	4.65	1.35	5.42	1.19	-3.59	**<0.001*****	0.61

Table 10: Statistics for the comparative analysis of the Stable Diffusion tool with generative.fashion on usefulness, ease of use, and satisfaction. Levels of statistical significance are denoted as follows: $**p < 0.01, ***p < 0.001$.

	Stable Diffusion		**generative.fashion**		**t-test**		
	M	**SD**	**M**	**SD**	**t(38)**	**p**	**d**
Quantity	5.49	1.32	5.87	1.13	-1.53	0.133	0.314
Quality	4.79	1.59	5.51	1.39	-2.34	0.25	0.480
Diversity	5.21	1.56	5.18	1.47	0.08	0.934	0.017
Pride	5.13	1.45	5.56	1.29	-1.77	**0.084+**	0.317
Meaningful Contribution	4.46	1.70	4.95	1.64	-1.62	0.113	0.292
Control	3.90	1.68	4.85	1.60	-3.04	**0.004****	0.578
Conceptual Realization	4.00	1.86	4.67	1.66	-2.18	**0.036***	0.378

Table 11: Statistics for the comparative analysis of the Stable Diffusion tool with generative.fashion on the DSEQ questions. Levels of statistical significance are denoted as follows: $+p < 0.10, *p < 0.05, **p < 0.01$.

A.6 Supplementary Figures

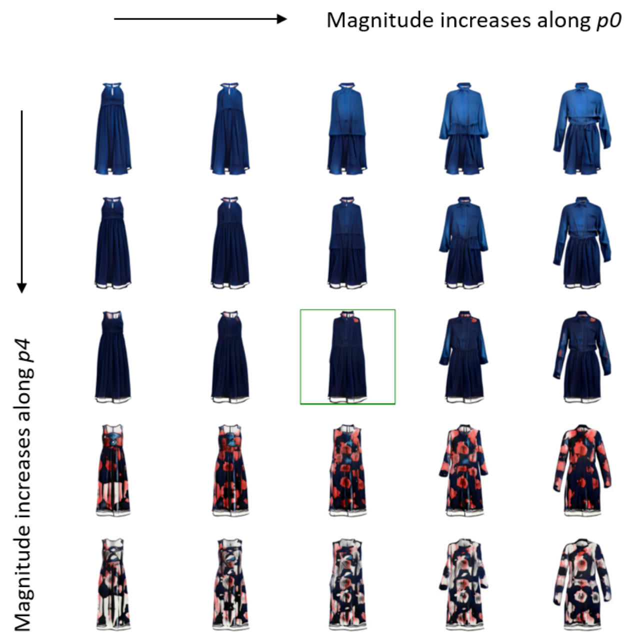

Figure A1: Examples of interpolating images simultaneously along two meaningful directions in the latent space (sleeve and pattern) found using PCA. The image in the green box shows the original image with 0 magnitude.

Figure A2: Examples of meaningful principle components p found in the latent space by PCA. (Dim k, Layers a-b) represents the k'th principle component applied in layers a to b in the 14-layer synthesis network. The images in green boxes are the original images with 0 magnitude. For each p, we show the result of two images, one in-sample and another out-of-sample.