

Exploring Normative Behavior in Generative AI: A Case Study of Text-to-Image Systems Using Stable Diffusion

Exploration des Comportements Normatifs dans l'IA Générative : Une Étude de Cas des Systèmes Texte-to-Image Utilisant Stable Diffusion

Michelle Melina Dutoit

LMU Munich

Munich, Germany

Sorbonne Université, CNRS, Institut des Systèmes

Intelligents et de Robotique, ISIR

Paris, France

michelle.dutoit@ifi.lmu.de

Baptiste Caramiaux

Sorbonne Université, CNRS, Institut des Systèmes

Intelligents et de Robotique, ISIR

Paris, France

baptiste.caramiaux@sorbonne-universite.fr

Abstract

Generative AI systems embed and transmit norms, raising critical questions for creative practice as they increasingly shape cultural production. This paper examines how creative practitioners perceive and experience normativity when using a text-to-image (T2I) system based on Stable Diffusion. We asked artists and designers to create illustrations inspired by two media sources with different levels of specificity, and analyzed their process through a mixed method design combining topic analysis, questionnaires, and semi-structured interviews. Results show that the studied T2I system often simplified the topics involved, particularly in low-abstraction tasks, and that creative practitioners perceived normativity through stereotypical outputs, invariant patterns, and unsolicited omissions or additions of details. These tendencies generated a sense of disempowerment, as creative practitioners had to adapt their creative process and linguistic choices to the system's constraints. We conclude by discussing how transparency and stronger user control could mitigate these effects to make this technology more empowering for creatives and artists.

Résumé

Les systèmes d'IA générative intègrent et transmettent des normes, soulevant des questions critiques pour les pratiques créatives. Cet article examine comment les praticiens créatifs perçoivent le comportement normatif des systèmes texte-image (T2I) basés sur Stable Diffusion. Dans une étude expérimentale, des artistes et des designers créent des illustrations inspirées de deux sources médiatiques présentant des niveaux de spécificité différents. Une analyse mixte combinant analyse des entretiens semi-directifs, des questionnaires et des images générées montrent que le système T2I étudié simplifiait souvent les sujets abordés, en particulier dans les tâches peu abstraites, et que les praticiens créatifs percevaient la normativité à travers des sorties stéréotypées, des motifs invariants, ainsi que

des omissions ou ajouts non sollicités de détails. Ces tendances ont engendré un sentiment de *desempowerment*, les praticiens devant adapter leur processus créatif et leurs choix linguistiques aux contraintes du système. Nous concluons en discutant de la manière dont la transparence et un meilleur contrôle utilisateur pourraient atténuer ces effets, afin de rendre cette technologie plus émancipatrice pour les créatifs et les artistes.

CCS Concepts

• Human-centered computing → Human computer interaction (HCI); • Applied computing → Arts and humanities; • Computing methodologies → Artificial intelligence.

Keywords

Normativity, Text-to-Image, Generative AI, Creative Practitioners

Mots clés

Normativité, IA Générative, Pratiques Créatives

ACM Reference Format:

Michelle Melina Dutoit and Baptiste Caramiaux. 2026. Exploring Normative Behavior in Generative AI: A Case Study of Text-to-Image Systems Using Stable Diffusion. In *Proceedings of the 37th Conference on l'Interaction Humain-Machine (IHM '26)*, October 19–23, 2026, Villeneuve-d'Ascq, France. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/3845948.3845964>

1 Introduction

With the rapid proliferation of AI-generated images, the question of whether these cultural artifacts reproduce normative views has become increasingly urgent. Generative AI systems tend to reduce countries and cultures to stereotypical representations¹, which is particularly problematic when such images are used to document or reinterpret historical events². Norms in visual representations shape what is considered acceptable, determine who is represented and made visible, and influence which perspectives are privileged or marginalized. Better understanding these normative behaviors, and how they affect the use of these systems by creative practitioners, is



This work is licensed under a Creative Commons Attribution 4.0 International License. IHM '26, Villeneuve-d'Ascq, France

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2965-2/2026/10

<https://doi.org/10.1145/3845948.3845964>

¹<https://restofworld.org/2023/ai-image-stereotypes/>

²<https://petapixel.com/2023/05/15/photojournalist-controversially-turns-to-ai-to-illustrate-inaccessible-stories/>

therefore emerging as a central concern in HCI research on creative practices with generative AI (genAI) [11, 15, 34, 38, 39, 49].

In this context, recent work has shown that T2I systems (Dall-E, Imagen, Midjourney, Stable Diffusion, Parti) misrepresent cultures and reinforce Western-centric norms [2, 11], as well as identities. As an example, these generative T2I systems inadequately represent South Asian culture by failing to generate culturally specific subjects, amplifying hegemonic defaults, and perpetuating stereotypical tropes [39]. Ethnographic work in Bangladesh further illustrates how such systems can limit creative ideation and yield distorted or inaccurate representations for local artists [34]. More broadly, generative AI technologies constrain representational diversity and reproduce normative identities and narratives [49], enacting what Gillespie describes as ‘politics of visibility’ [15]. Collectively, these studies provide community-driven investigations of T2I systems, documenting how they embed biases and reproduce normative patterns that render certain identities and narratives invisible, with harmful consequences for the affected communities. Building on this work, we examine the normative behavior of T2I systems (in particular a diffusion-based system) from the perspective of creative practitioners who are new to these tools and not necessarily targeted by the norms they convey. Our goal is to understand the processes through which these norms manifest themselves and their impact on practitioners.

This work is therefore part of a larger body of research that has examined how artists develop strategies to integrate these systems into their workflows [7, 28, 32, 40, 44]. With the growing adoption of generative AI, and T2I systems in particular, creative practices are beginning to evolve in significant ways [7, 44]. A practice of T2I systems starts to emerge, where creatives design reusable prompt templates to adapt visual styles [7]. In order to circumvent normative behaviors, certain artists adjust keywords to better align outputs with expectations [38], or refine prompts iteratively to steer the model toward desired outcomes [49]. Yet these studies often center on contexts where intentions are particularly high—either among experienced practitioners or within communities misrepresented by existing models. As Mahdavi Goloujeh et al. [32] show, users with highly specific intentions frequently struggle to obtain satisfactory results, while more exploratory engagement can mitigate these challenges. This suggests that the specificity of a task plays an important role in shaping how creative practitioners experience normative behaviors in T2I systems and the strategies they adopt to navigate or resist them, an aspect that remains underexplored.

In this article, we address this gap by investigating how creative practitioners engaging with a text-to-image (T2I) system perceive and navigate the norms embedded within it across tasks that vary in specificity, understood as the level of detail available within the task. Our research questions are twofold:

- **(RQ1)** How does the level of task specificity shape the creative practitioners’ experience of normative behavior of the T2I system?
- **(RQ2)** How do creative practitioners perceive the normativity of the T2I system, and how does it affect them?

To address these research questions, we conducted an experimental study with 14 creative practitioners, located in Western Europe. We

asked them to use Stable Diffusion, a widely used T2I system, to create illustrations of a textual story taken from a newspaper and a soundscape. These two media were used to create two qualitatively distinct levels of specificity for the requested illustration task. The textual story provides specific details about the scene to be illustrated while the Sound Task leaves more room for interpretation. We used a mixed quantitative and qualitative method to address the research questions. To examine RQ1, which concerns the influence of task specificity on the perception of the system’s normativity, we carried out a comparative study combining a questionnaire with an analysis of the topics present in the prompts and the images generated. We conducted a series of semi-structured interviews to address RQ2, to investigate the impact of these perceptions on their creative behavior. Our results showed that it was easier to control the system in the less specific task, and that the system involved a more radical simplification of the themes addressed in the more specific task. Additionally, we found that participants perceived the normativity derived from the Stable Diffusion-based T2I model through recurring, stereotyped patterns, and added or omitted elements independently of the artist’s indications. The friction encountered by artists in using the system led to a feeling of disempowerment on the part of participants, who tended to abandon elements that seemed impossible to produce, focused more on aesthetics than content, and formed compromises. Overall, our work contributes to the question of normativity in visual generative AI from a conceptual and practical point of view. It is important to note that this study is not a systematic evaluation of the T2I model capabilities and its norms, but an exploration of normative behavior of genAI, as we only explored one T2I model (Stable Diffusion) and only conducted our study in Western Europe. Our goal is to provide insights that will help identify directions for future research on this topic in HCI.

The structure of this paper is as follows: We begin by reviewing the state of the art, including prior research on T2I systems in the creative domain, and normativity in AI. Then, we outline the methodology employed in our study and detail the approach to data analysis. The next section presents the study’s findings, focusing on participants’ perception of normativity, and its influence on them. Finally, we discuss the implications of these findings for the community.

2 State of the Art

In this section, we first review background work on creative practices with GenAI-based T2I systems, followed by prior research that addresses the concept of normativity in these systems.

2.1 Background on T2I for creative practice

As part of research into AI practices in creative and artistic contexts [3, 6, 13, 19, 25], recent research has looked in particular at practices surrounding the use of AI-powered T2I systems in creative and artistic fields [7, 33, 40, 44, 46]. Artists who are using T2I systems do so mainly for ideation and production [40]. They develop strategies to achieve the desired results, such as creating a prompt templates and adapting their language to match their expectations [7, 32]. Analysis of human-created prompts led researchers to propose guidelines for designing them [28].

Recent research also reported shortcomings in the use of T2I in creative tasks. One shortcoming is the difficulty of expressing thoughts in words, which is necessary to create a prompt [33]. More broadly, not all artists may feel comfortable using language as their primary mode of expression. Another limitation of using T2I systems is the notion of “concept entanglement”, which refers to the system’s inability to represent specific concepts independently of other concepts [44]. McCormack et al. [33] also refer to this phenomenon, associating it with the fact that the conceptual understanding for this type of technology is not comparable to that of humans, creating misunderstandings in the use of these systems.

Finally, recent work by Mahdavi Goloujeh et al. [32] has shown that the task has an impact. Users with highly specific intentions often struggle to achieve desired outcomes, whereas more exploratory engagement with the system can mitigate these difficulties.

2.2 Normativity in T2I systems

The concept of norms is related to the moral judgment of what is conceived as good or bad to do. A recent study on language models showed that Language Models (LMs) retain norms by clustering texts categorized as morally good or bad together [45]. In another study, Gillespie [15] demonstrates that current text-based generative AI tools do tend to replicate normative identities and narratives, and when cultural categories and identities were not mentioned, non-normative alternatives rarely were produced.

In this paper, we are specifically interested in normative behavior of T2I systems. In a recent essay, Donnarumma [11], discusses the norms imbued in these systems and relate them to their opaque nature. Because AI technologies are very costly resource-wise, and rarely open [50], only a small amount of companies can afford their research and development. Therefore these companies have the power to define the AI aesthetics that grow culturally dominant and become the norm, what the author called corporate AI aesthetics. Recent studies have then addressed the question of normativity in T2I systems through field work and empirical studies, raising concerns about the biased gaze that these systems provide. For instance, Qadri et al. showed how T2I can reproduce an outsider gaze for South Asian cultures and they showed how engaging with communities allowed for designing new assessment methods of these systems [39]. Similar shortcomings have been observed in Bangladesh, where T2I systems struggled to capture local cultural contexts [34]. These skewed perspectives towards Western culture is not the only normative trends observed in these systems. Such systems tend to output negative, reductive representations of people with disabilities which replicate and result in real-world harm [31]. They suggest to focus on creating datasets with better representation or strive to use Machine Learning (ML) techniques that rely on less, and better curated, data. Finally, queer people are also subject to misrepresentation in T2I systems which embed normative values such as hyper-positivity and limiting the representation of queer sexuality [49].

Collectively, these previous studies tend to illustrate the values and culture conveyed by these systems, but less so the processes through which these normative behaviors operate. Furthermore, little research has been done on strategies for circumventing these behaviors. In Taylor et al. [49], the authors showed that queer

artists in their study often repeated prompts to test the model, devised workarounds for content moderation, and iteratively refined prompts to achieve their objectives. In another recent study, Qadri et al. [38] examined how artists appropriate T2I systems for non-Western cultural expression. They found that participants developed strategies to circumvent the normative constraints of these systems, such as testing with multilingual prompts. Artists also looked for disrupting stereotypical associations by adding or removing keywords in an A/B fashion, and attempted to bias the model through the deliberate introduction of their own cultural norms.

3 Methods

3.1 Participants

We recruited 14 creative practitioners, including one visual artist, five graphic designers and four illustrators, one 3D graphics designer, one communication designer, and two UX designers, who had a background in illustration. Eight participants identified as female, one as non-binary and five as male. The mean age was 32 years (SD=11.5). They were recruited from the authors’ networks, and through calls advertised on social media. The participants have never used, or only tried, generative T2I systems before. The rationale behind this was to avoid having participants who had already learned to overcome some of the limitations of generative T2I systems (as shown, for instance, in [7]), in particular their potential normative impacts.

3.2 Tasks

Participants were asked to complete two tasks. In each task, they had to create an illustration from a given source media, using the T2I system provided (see Section 3.3). In the Story Task, participants were asked to create an image inspired by a written newspaper article, which is disclosed in the Appendix B. The story describes a man in his forties and his grandmother visiting the national parks of the USA. We chose this story because it represents a grandmother-grandson relationship with both of older age, which is not often depicted in illustrations and therefore is difficult to generate with generative AI.

In the Sound Task, the participants were asked to create an image inspired by 30 seconds of audio. The stimulus was a soundscape where footsteps can be heard on the ground, with surrounding birds and wind. The Sound Task therefore describes a situation similar to the Story Task, i.e. people hiking in the great outdoors. However, the Sound Task is more abstract than the Story Task, in that the characters and location are left to the participants’ choice.

3.3 Setup

The study was conducted in two different contexts: in person, where the first author conducted the study on our institution’s premises; remote, where the first author conducted the study using a video conferencing tool.

3.3.1 In Person. The participants were given a laptop with the AUTOMATIC1111 Stable Diffusion Web UI [4], which can be seen in Figure 1. We chose to use Stable Diffusion because it allowed to be used locally on the computer. This allowed to better handle the

time for image generation, personalizing the interface and make it possible to not have a restriction of how many prompts the participants can use. We customized the Stable Diffusion Web UI by only displaying the functionality of writing a prompt in a text field, clicking the button to generate, and seeing the generated picture. We chose the model Stable Diffusion XL ([37], downloaded April 5, 2024) since during the time of the project it was the latest and most downloaded Stable Diffusion model on Hugging Face. The questionnaire was created with SoSci Survey [23], which follows German privacy restrictions, and was made available online³. The sound snippet was played on the laptop given to the participants, and the article was given to them as a paper copy.

3.3.2 Remote. For recruitment reasons, we also conducted the study remotely. Gradio [1], used to implement the Stable Diffusion Web UI, makes it possible to host an online application by creating a temporary link⁴. This allowed the application to be shared with participants. However, the data remained logged on our local computer. The online installation then included the Stable Diffusion interface configured as described above. The other study resources, i.e. the information letter, consent form, Zoom link, questionnaire, sound and article, were e-mailed to the participant prior to the study. The online meeting was hosted by Zoom, using our institution's subscription.

3.4 Procedure

The study followed a within-participant design and was reviewed and approved by the institution's Research Ethics Committee. The study had 3 phases: Introduction, Tasks with Questionnaire, and Interviews. Overall the study lasted about one hour. In the first phase, the participants were welcomed to the study and given the information letter and the consent form for the study.

In the second phase, the participants were asked to perform two tasks, which are described above. First, the participants were given the Sound Task, and then the Story Task. We chose this order, rather than counterbalancing it, because we did not want the details of the story to be illustrated to influence the illustration of the soundscape, given that this soundscape was designed to convey the image of people walking in nature. The goal was to give participants the freedom to interpret what they heard in an exploratory manner, at the expense of a risk of an order effect between the two tasks.

After each task, they were asked to fill out a questionnaire about their experience, process and evaluation of the tool. The questionnaire items are disclosed in the Appendix C.1. Some of the questions of the questionnaire were taken from the Creativity Support Index [8], which measures the expressiveness and exploration of the tool, and from the AttrakDiff standardized questionnaire [18], an established UX evaluation tool. Additionally, we added ad-hoc questions about normativity based on our literature review.

In the third phase, an interview was conducted, which addressed how they perceived the normativity in the system, specifically by asking how they evaluated their freedom of expression, assistance, and neutrality of the tool. The interview questions are reported in Appendix C.2. During the interview, the goal was to initiate a reflection about their experience and if they perceived a normative

basis from the system. If it was unclear to the participant what was meant with normative basis, it was clarified with "if the system interpreted the prompt on the basis of certain norms."

3.5 Data Collection and Analysis

During the study, we collected the responses to the questionnaires, the prompts used by the participants, the images generated by Stable Diffusion from the prompts, and finally the interview recordings were collected for analysis. We followed a mixed methods approach with an explanatory sequential design, analyzing the quantitative data first and following up with the qualitative data to help explain the quantitative results [9]. We analyzed this data as follows:

- *Questionnaire Data.* The questionnaire data was analyzed through first and second order statistics to compare the participant's experience, process and evaluation of the tool between the Sound and the Story Tasks (RQ1). Specifically, we applied an ANOVA test on the questionnaire results with the (1) the task (Sound, Story), and (2) the questions answered. The test was applied after normalizing the collected data.
- *Prompts and Generated Images Data.* We also wanted to analyze what participants produced during the tasks. There are no standard methods to analyze prompts and images. Here we chose to use topic analysis as a way to assess the topics involved in the production of the illustration. Specifically, we conducted the following analysis:
 - *Analyzing prompts.* We used BERTopic [16], a topic modeling technique based on Large Language Models (LLMs) that creates clusters of topics from text descriptions (note that this model was already used to analyze prompts in T2I in [44]). A dictionary was created with each word that was used in the prompts. This was done to normalize the amount of words used by the participants, as these varied with some using more words and prompts than others. We then used BERTopic to organize them into topics. The model automatically groups similar words into coherent topics.
 - *Analyzing images.* We wanted to reproduce the topic analysis on images, so that we could compare inputs (prompts) and outputs (images). To this end, we borrowed the method of Luccioni et al. [29], who used a caption generation model to analyze the content of generated images as a method to explore social biases in T2I systems. They generated captions from images with the generative models ViT GPT-2 Model and BLIP, and computed the likelihood that a caption uses gender-marked words. Inspired by this work, in this paper, we used two types of model: BLIP [24] and Florence-2 [51], to enhance the interpretative capacity of our method. Both models generated a textual description of each image. We then used the same analysis process as described above, generating overarching topics of the words used in the captions with BERTopic.

The results from the analysis were applied twofold: first, as a means to compare the topics involved between prompts and images, and second, to assess the tasks by evaluating both the quantity and nature of themes across the different tasks.

³via www.sosicisurvey.de

⁴<https://www.gradio.app/guides/sharing-your-app>

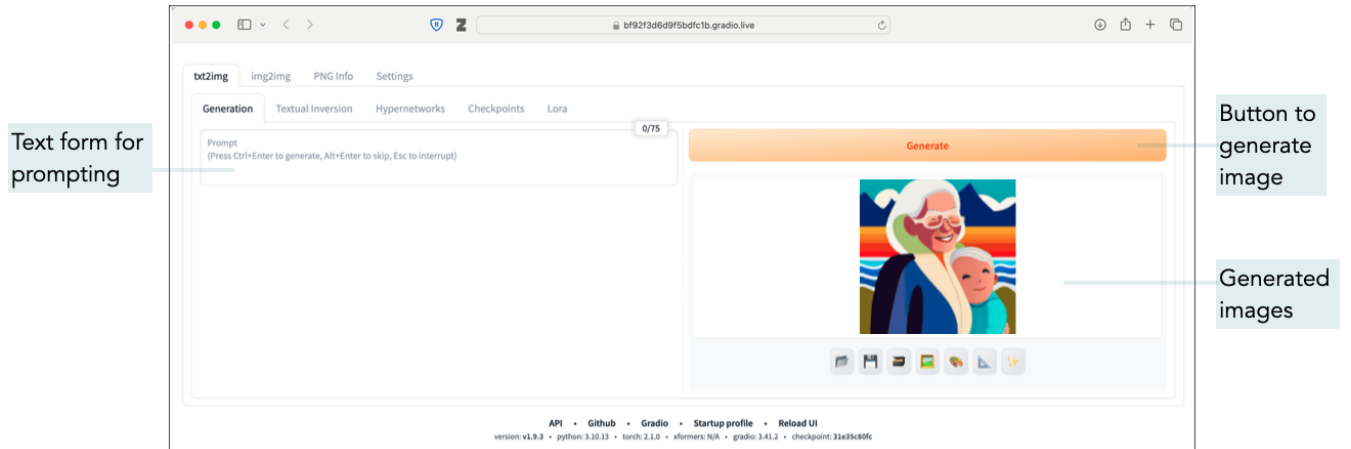


Figure 1: The customized AUTOMATIC1111 Stable Diffusion Web UI.

- *Interview Data.* The recordings were transcribed and translated, as the interviews were conducted in English, German and French. For this, we used Whisper from OpenAI locally on a secure computer, and reread the transcripts to ensure its validity. Our research team is fluent in French, English and German, which allowed us to ensure an accuracy of the translations. We then analyzed the transcripts using an inductive thematic analysis, as described in [5]. First, both researchers familiarized themselves with the data by reading it several times and adding notes. Then, both researchers independently created codes, which were a summary of a part of the data, or a description of the meaning of that part of the data. These codes were then grouped together to create a common code book. The two researchers then organized the codes independently into themes and sub-themes. These themes were then discussed, refined or eliminated until a consensus was reached.

4 Results

In this section, we present the results of the experimental study. We start by reporting descriptive statistics about the tasks. Then we present the results of the analysis of the prompts and illustrations produced, which provide quantitative accounts of how participants engaged with the tasks. Finally, we present the results of the thematic analysis, which enables us to interpret normativity through participants' verbalized experiences.

4.1 Questionnaire analysis

Figure 2 and 3 depict some example images that were generated from both tasks. From the questionnaire, we found that participants self-reported that the images generated during the Story Task aligned less with the expectations than during the Sound Task ($F(1, 20) = 2.1, p < 0.05$) and participants felt less creative during the Story Task ($F(1, 20) = 2.2, p < 0.05$). This provides insight into the system's normative behavior by showing how its responses differed between the Sound Task, which is more open-ended and

less anchored in visual norms, and the Story Task, which required depicting a motif that explicitly challenged those norms.

Takeaway: The more specific task leads to creating images that are less aligned with expectations, resulting in a feeling of being less creative.

4.2 Prompt and generated images analysis

We analyze the topical complexity of both the prompts and the corresponding generated images to (1) determine whether participants created prompts with higher topical complexity in one task compared to another and (2) to evaluate whether the images generated reflect a level of topical complexity comparable to that of the prompts used to generate them. To achieve these goals, we used computational topic analysis (described in detail in Section 3.5). Prompts and generated image captions were transformed into a dictionary with all of the words used in the prompts, which was then used in the topic analysis (see Methods). Table 1 reports the number of words and topics obtained.

Task	Type	# words	# topics
Sound	Prompts	382	7
	Captions (using BLIP)	221	5
	Captions (using Florence)	199	6
Story	Prompts	507	16
	Captions (using BLIP)	283	6
	Captions (using Florence)	260	4

Table 1: Number of Words and generated Topics per Task and Type

First, we found that participants used more different words in the Story Task than in the Sound Task (507 and 382, respectively i.e. +32.7%). We also found a difference between the amount of topics involved in the Story Task and in the Sound Task (16 and 7, respectively, i.e. +128.6%). The difference in number of topics may be linked to the fact that participants used more words in



Figure 2: Examples of Generated Images during the Sound Task.



Figure 3: Examples of Generated Images during the Story Task.

the second task. However, given that the percentage increase in the number of topics is greater than the increase in the number of words, this suggests an increase in topical complexity.

Second, our analysis revealed a reduction in the number of words between the prompts and the captions generated by *BLIP* (-42.1% in the Sound Task and -44.2% in the Story Task) and *Florence* (-47.9% in the Sound Task and -48.7% in the Story Task). More importantly, we observed a decrease in the number of topics identified in the captions compared to the prompts. While this reduction in topics was relatively modest in the Sound Task (from 7 to 5 topics in *BLIP*-generated captions and from 7 to 6 topics in *Florence*-generated captions), the decrease was more significant in the Story Task, with topics reducing from 16 to 6 in *BLIP*-generated captions and from 16 to 4 in *Florence*-generated captions.

Figures 4 and 5 depict the topics extracted from prompts and both *BLIP*- and *Florence*-generated captions for both tasks. The topics extracted from the captions generated for the images in the Sound Task by *BLIP* and *Florence* are overlapping topics extracted from prompts (Figure 4), indicating that the system does not operate simplification or significant change of topics between the prompts and the generated images. The topics extracted from the captions generated for the Story Task images by *BLIP* and *Florence* showed a substantial reduction compared to the prompts (Figure 5). Notably, topics that provide more specific descriptions of the images, such as adjectives, weather, and actions (e.g., viewing and traveling), were absent. This suggests that participants faced constraints in

generating detailed and specific prompts, as the resulting images appeared overly generalized.

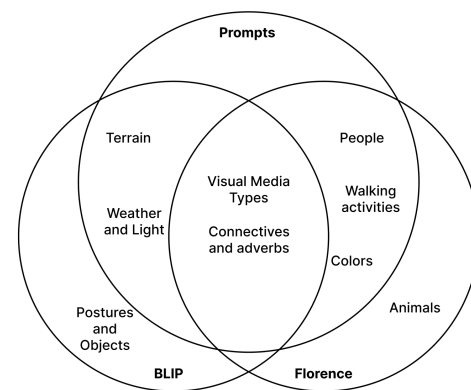


Figure 4: Sound Task Topic Analysis Results with Overlaps.

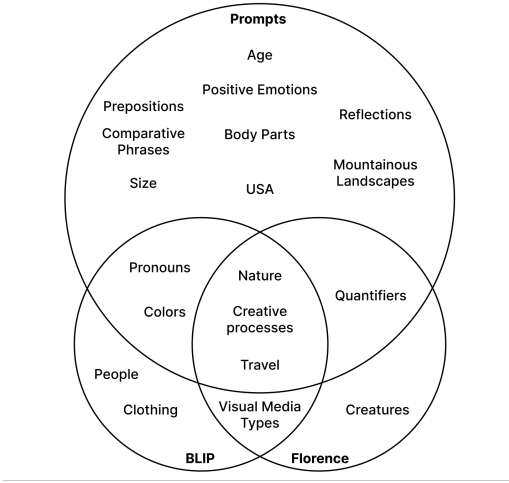


Figure 5: Story Task Topic Analysis Results with Overlaps.

Takeaways: For tasks involving non-abstract concepts, such as the situations and characters required in the Story Task, the Stable Diffusion model tends to simplify the intended representations, producing visual outputs that are less detailed than the textual inputs. The lower expectations reported for the Story Task in previous section can be understood, at least in part, as a response to this simplification.

4.3 Interviews

Previous analyses offered insights into the system’s normative behavior by examining how participants engaged with the tasks and how the system responded. In this section, we present findings from a thematic analysis, which provides a deeper interpretation of normativity through participants’ verbalized experiences. Specifically, two key themes emerged from the interviews: participants’ perceptions of the system’s norms and how they impact their behavior.

4.3.1 Perception of the System’s Norms. Our findings indicate that participants noticed stereotypical behavior exhibited by the system, identified normative patterns through its invariant responses, and attributed the system’s normative behavior to omitted details in prompts or additional details introduced by the system.

Normativity in stereotypical representations. First, participants identified the system’s normative behavior as coming from stereotyped, biased imagery. P11 pointed out, she could see the norms in the way an old lady was pictured, or how a mountain is generated, which “was very much like a stereotype of what a mountain is, or what a stereotype of what the grandma is. Which is norm.”. In addition, P6 highlighted that the system “conveys the western culture norms, middle class” and has “like a commercial, low key visual culture.”. This observation is in line with previous work that has studied the biases and stereotypes inherent in generative models, such as Luccioni et al. [29] who showed the biases linking gender and profession, or Qadri et al. [39] who showed the Western view on South Asian culture conveyed by these systems.

Normativity in invariant patterns. The participants saw repeated patterns as another manifestation of the system’s normative behavior. Several participants, including P3, P4, and P7, observed that all the images shared certain commonalities, while having difficulties to articulate these patterns as P3 explained: “I can’t say why. But simply because [with] all the pictures, there is a certain something that every picture had in common.”. P4 described that he thought they were similar patterns in the images, in terms of the composition, layout, and perspective: “I have the feeling that portraits always look very similar from the layout. And other scenes too, there are often very classic, central one third perspective.”. P14 had the strategy of probing perceived patterns by repeating the same words in prompts and inspecting invariants in the images generated: “[I] repeated often the same prompt and looked at whether something changed and most of the things stayed the same and are probably these conventional things.”. A similar strategy was highlighted by Salvaggio [43] or Gillespie [15], who explored the emergence of patterns in T2I systems through repetitive prompting. According to Salvaggio, identifying these invariants can provide insights into the underlying training data, where strengths in generation correlate with prevalent elements in the training set, and weaknesses reflect areas where the training data is sparse. Overall, invariants convey norms.

Normativity in the omitted details. We found that participants perceive normativity in the system’s behavior through details specified in the prompts but ignored by the model. P14 noted that even when she included specific details in her prompts, these elements were consistently excluded from the generated images. Furthermore, attempts to elaborate on details and use more descriptive or complex prompts proved ineffective, as reported by P2, P3, P5, and P14. For example, P2 explained “And then the prompt became more and more complex and longer. And the more complex the prompt became, the more often things were left out.”. The omitted details concerned the male character who was often omitted and mainly represented as having the same age as the older woman, as P12 described “I then also included, for example, an adult man and there was practically no second person included”. The details in the prompt guiding the perspective of the shot were also omitted, as it is often shown in the third person: “I was trying to have the point of view of the dashboard from the backseat with the faces illustrated in the mirrors. (...) I didn’t get any rear view mirrors” (P11). Finally the details related to the visual style were also omitted, as P11 described: “The division of the layout, the line trace, like the line style, even the colors at some point, I couldn’t really find a way to tell it to create black and white for a specific area and colors for another. All of the details I was trying to put in the second prompt, I felt like they were dismissed”.

The topic analysis (reported in Section 4.1.2) supports the participants’ perception and shows that the specific topics left out are, for example “Age and Generations”, including words like “old”, “grandma”, and “aged”, and the topic “Reflections”, including words like “mirror” or “reflection”, illustrating the omitted perspectives.

P5 argued that this omission of details highlights the system’s lack of neutrality: “in my opinion my prompt was not neutral, because my prompts were generated in my head and the interpretation

of the tool is also not neutral, because it either intentionally or unintentionally left some details out.” According to P5, a truly neutral tool would faithfully generate all elements specified in the prompt.

Normativity in the added details. Consistent with the omission of details, participants theorized that the system performed better with less specified prompts, as too many details in prompts were often ignored. P8 and P12 observed that when fewer details were included, the system generated imagery that conformed more closely to normative representations: “I didn’t write that the grandma has to have white hair. I just said an old woman wearing a flower dress and glasses. But for example, white hair was probably the cliché picture. And the other man, funnily enough, although I didn’t describe an age or anything else, he was just as old as her.” (P12). Other participants highlighted specific assumptions related to age and race. For instance, the man depicted alongside the grandmother was consistently rendered as being either of a similar age or the age of a child, and the grandmother was exclusively portrayed as being of Caucasian origin. P8 for example commented “The interpretation it takes it’s not so precise, although I already tell him it’s a 40 years old guy, it makes him look really old, so then I stop caring about it.” For P2, using less specific prompts became a deliberate strategy to probe the system’s normative biases: “In the second task, I actually also tried to be super unspecific, so that I might also overstrain the AI, because what comes out is less expected, and that’s what worked.” He distanced himself from the concept of generating a grandmother and grandson, opting instead to create abstract and non-normative content.

4.3.2 Impacts of System’s Normativity. The aim of this research was to understand the impact of system normativity on participants and how it might govern their behavior. Our results suggest that there is a sense of disempowerment among participants, who describe feelings of powerlessness and wanting to surrender. Participants were forced to compromise on their initial concept and try to find a satisfactory outcome within the norms.

Feeling of disempowerment. Some participants reported feelings of powerlessness in controlling the system’s outputs. P14 said: “So you could say that I assisted the tool, but I don’t know if the tool needed it.” In response to this perceived lack of control, a common strategy was to give up efforts to influence uncontrollable aspects: “I would rather say I gave up. The system didn’t understand it. Or it hasn’t learned enough yet. So if you look at from the training data, then it does not have the data at all. So it doesn’t matter how I communicate with the system, because the data is just not there.” (P1). The participant attributed the system’s limitations to its inability to generate abstract content, instead producing only conventional layouts. P8 also described giving up on aspects she could not control: “So I start using some words once I find out, for example like its ability is not so good when generating human faces and I don’t care about it, so I start typing something else like focusing on the image for example like the styles and the big picture that I want to present, for example like grandma and grandson they’re working together, but I don’t focus on the face”.

This feeling of disempowerment was stronger in the more specific task (Story Task), which was supported by the questionnaire results revealing that the images generated in this task were less in

line with participants’ expectations than in the Sound Task. Similarly, the drastic reduction in the number of subjects in the analysis of the subjects in the Story Task, from 16 subjects to 6 (BLIP) and 4 (Florence), shows a limitation of the participants in what they were able to generate. This suggests that users may forgo efforts to create non-normative content due to the inherent challenges in achieving such outputs.

Making compromises. A common strategy observed among participants involved compromising between their original concepts and the system’s capabilities, as P2 explained “So I then thought about what is, so to speak, expectable, what the program does and how can I adapt my concept so that I get an outcome, that there is a compromise between what I imagine and what the program does.” These participants appeared to expect normative content from the system and adjusted their concepts to align with what they believed the system could produce effectively. Similarly, P11 employed a strategy of compromise by lowering her expectations: “And then when I started using the program and realized the types of images that I was receiving, it made me re-adjust my expectations. And I mean, watching the prompt being generated, (...) I always had another image in mind [than] the one that was generated. And I was trying to expect something.” Participants adjusted their prompt language by relying on conventional descriptions that they believed the system would better comprehend, as P10 put it: “conventional pictures or descriptions, for example, (...) family photos in the regular shot, the family members looking in the camera”. Or participants adjusted the framing of outputs to conceal the system’s limitations. For instance, P8, dissatisfied with the system’s rendering of faces, opted to either zoom out or focus closely on specific features, thereby achieving outputs that met her expectations more effectively.

A different compromise was also explored by abandoning any pre-existing concept and exploring what the system could generate, as P2 described: “And for the second time it was rather that I realized that it won’t illustrate what I want anyway so then I’d rather be surprised and see if anything cool comes out”, explaining that he challenged the T2I system with abstract and unconventional prompts.

This compromise led to an iterative process, described by P1, P4, P8, P10, and P12. As P10 explains: “From the images that it proposed to me, I tried to make them evolve, but in the same direction. In fact, it proposed images of a road with a landscape. And so I tried to make this image evolve, but still looking at the road with the landscape.”. Participants noted that the system influenced them in a way that they tended to stay within the concept suggested by the first image and simply iterate on it.

Takeaway: The system’s normative behavior is perceived through invariant patterns, stereotypical outputs, and the omission or addition of details. These behaviors have concrete effects, including eliciting feelings of disempowerment among users and forcing them to make compromises in both visual creation and the language employed to interact with the system.

5 Discussion

In this paper we conducted a mixed methods study investigating the perception and impact of the normativity of generative AI tools,

specifically Stable Diffusion. We structure our discussion in three sections, discussing the impact of the tool's default behavior, the improved transparency of the prompt, and the creatives' sense of powerlessness. Finally, we present the limitations of our study.

5.1 Beyond default standpoints

Our study examined how creative professionals - users specifically targeted by T2I systems — experienced normative depictions provided by such systems. These normative behaviors were perceived as stereotypical representations. In practice, normative behaviors were perceived as stemming from the system autonomously deciding to omit or add details, which was further supported by the topic analysis revealing a higher reduction in topics during tasks involving narratives with context and character development. This reduction reflects a simplification of visual representation, favoring default depictions, or, as Salvaggio put it, repeating the unconscious assumptions that are embedded into human-assembled datasets [43].

As Gillespie emphasized, defaults are significant because, even when users can explore alternatives, they tend to adhere to default options [15]. These default options lead to the reproduction of limited perspectives, thereby restricting the visibility of alternative viewpoints and identities. The study of Mack et al. [31] supports this, as the participants found that the T2I models did not display adequate diversity by default. They highlighted that there is a necessity for diverse output by default, as prompt engineering is not obvious or possible for everyone. In the more general context of creativity-support tools, Li et al. [25] discuss the case of artists who use the default abstractions of the tools and are able to 'make do' with them, taking what is immediately available, but not being able to explore other possibilities.

Default options in generative systems often embody implicit cultural and political biases [11], effectively imposing specific standpoints on users. For example, they may reinforce narrow stereotypes, such as depicting relationships between older women and middle-aged men through a particular lens, or portraying grandmothers uniformly as short-haired Caucasian figures, sometimes accompanied by a cane. These biases are not only embedded in the technology itself but can also be exacerbated by artistic practices. Some artists, for instance, rely on standardized prompt templates in their workflows, as documented by Chang et al. [7], further perpetuating and amplifying these default assumptions.

Research recommendation. We argue that it is essential to advance research on the default behaviors of text-to-image generation systems and to explore strategies for moving beyond these default viewpoints. On the one hand, default perspectives arise as a byproduct of the stochastic nature of model inference. By leveraging, rather than averaging out, this stochasticity, it is possible to circumvent these inherent biases. A recent study by Qadri et al. [38] illustrates this approach, showing how artists iteratively repeated prompts to bypass culturally normative outputs. More broadly, it shows the need for further research on designing with the inherent uncertainty of these systems [14]. On the other hand, enabling deeper interaction with training data and learning procedures may help creatives push beyond default representational standpoints [12]. Techniques such as LoRA [20], Custom Diffusion [22],

or DreamBooth [42] allow diffusion models to be fine-tuned on small curated datasets and have been shown able to reshape the aesthetics of specific concepts. However, it remains unclear to what extent such fine-tuning methods can meaningfully counteract the socially embedded representations and biases present in large-scale pretrained models.

5.2 Towards prompt transparency

The system's deviations from participants' prompts made participants reflect on the decisions performed by the system. Specifically, the system often removed certain details from the textual descriptions while adding others without explanation. This led one participant to question: "Wouldn't a neutral generative system be one that transparently interprets textual descriptions without omissions or additions?" This comment highlights the importance of *prompt transparency*. The goal of such transparency is to enable relevant stakeholders to develop an accurate understanding of the system's capabilities, limitations, mechanisms, and methods for controlling its outputs [12]. It allows to increase the control the user has, as well as enable appropriate levels of trust which allows to accurately judge the benefits but also the shortcomings of AI output [27].

This aligns with prior research on explainability and transparency in AI [27, 47], which has emphasized documentation [35], interaction-focused approaches [26], counterfactual analysis [36], and black-box explanation tools [30, 41], as means to foster transparency. Relevant to prompt transparency, existing techniques such as LIME [41], SHAP [30], and neural integrated gradients [48] have been employed to enhance transparency in language models by highlighting significant tokens in the input that contribute to a given output. Extending these techniques to T2I models could help highlight omitted elements in prompts. However, a complementary approach is needed to explicitly identify and make transparent the details added by the system. One potential solution is the use of generated captions to highlight discrepancies between the input prompts and the outputs.

Research recommendation. Overall, transparency has primarily been investigated in high-stake application domain, and we believe that it needs to be better tackled in the creative domain because this domain shapes the cultural fabric of our societies. In a recent study, artists also discussed the benefit of transparency in fostering creative practice while informing about the appropriate uses of T2I [46]. Researching prompt transparency can enhance understanding of the limitations and capabilities of T2I models, enabling their more effective application and appropriation as creative tools, medium, and cultural instruments.

5.3 Disempowerment

The creatives who took part in our study often expressed frustration when using the system, and some felt powerless to control it. Mahdavi Goloujeh et al. who studied the prompt journey of active Midjourney users, also described frustration as a challenge in use [32]. When users had focused objectives and well-defined visions, they often experienced frustration when the system failed to generate important elements of their visions. The feeling of frustration and powerlessness is linked to a loss of user autonomy. We see this as an interesting contrast, as generative AI systems are often

marketed as empowering tools, enabling their users to perform a task much better and faster. For example, Stability AI claims “Stable Diffusion is a text-to-image model empowering billions of people to create stunning art within seconds.”⁵ In our study, the participants had a high level of visual literacy, and felt limited by what could be generated, perceived as normative. Our hypothesis is that they have the expertise to quickly detect conventional and stereotyped layouts, images and compositional choices. Naive users, who lack visual literacy, will be less aware of these limitations and are more likely to feel empowered by a T2I model. This hypothesis remains to be further researched.

Disempowerment through generative AI is not a topic much explored yet. Previous research did show a disempowerment of communities through the output of T2I models perpetuating negative stereotypes and narratives of marginalized groups [31, 39]. More generally, the feeling of disempowerment, and a loss of autonomy can create a *chilling effect*, as expressed by Jiang et al. [21], and dissuade creatives to be involved in the development of creative-centered AI tools. The “democratization of art” is negated through the harms these technologies bring to the creative community. Recent advancements that offer users greater control, such as the fine-tuning methods discussed above, represent a promising direction. However, these approaches also introduce new challenges, including the extensive tinkering needed to identify effective adjustments [10]. Emerging models such as Gemini 2.5 Flash Image (Nano Banana) similarly provide enhanced control over the generation process, yet they come with their own limitations: being proprietary, they restrict opportunities for further fine-tuning, hinder explainability, and limit transparency.

Research recommendation. This feeling of disempowerment goes against an important objective of system design, that a tool should empower the user to allow them to achieve their goals [25]. Li et al. argue that it is important to investigate how a tool influences the creative situation, beyond the usability and functionality. Therefore, an important research challenge is to investigate more closely the construct of disempowerment in genAI-based creative tool practice.

5.4 Limitations

Normativity is a challenging concept to research. This is because we all are participants in our own normative environments, which we are bound to [17]. To mitigate this and still get a satisfactory study design we focused on a very specific norm, which is the depiction of a grandmother-grandson relationship of an older grandson, which is not a normative depiction often found in illustrations and thus difficult to generate with generative AI. We believe that the study already provides useful insights, but a more thorough study involving a better control of the normative framework should be considered. In addition, we would like to highlight the exploratory nature of our work. Our study has limitations in investigating only one T2I model (Stable Diffusion), having a limited sample of participants, and being conducted only in Western Europe. We do not position our study as a definitive commentary on this or other T2I models, but to raise awareness that with the use of T2I models,

the generated imagery will be limited by a normative basis. Another limitation that we would like to highlight are the models that we used for the topic analysis and image captioning have their own biases and own normative grounds, which we are not able to control for. We tried to compensate these by using two models for the image captioning. Also it has to be noted, that the images simplified the prompts, and also the generated captions were a simplification of the images. So while the captions captured the topics of the images, they also contributed to the overall simplification. Another approach to the topic analysis that could be considered is an analysis by multiple people who classify the topics they see, and analyzing these. Further research is needed to address these limitations and advance the research on normativity in AI-powered text-to-image generative models.

6 Conclusion

In this paper, we researched the questions if normativity of generative AI systems, especially in Stable Diffusion, is perceived by creative practitioners, and how it impacts them, depending on the tasks’ specificity. We designed a study where creative practitioners were commissioned to create images with generative AI, with the inspiration of two different media of different levels of abstraction. We found that participants perceived the system’s normativity through invariant patterns, stereotypical output, and the omission or addition of details by the system. Participants were impacted by feelings of disempowerment and making compromises to achieve a better result. Through our topic analysis, in which we analyzed the topics used in each task and generated in the images, we found that tasks involving non-abstract concepts, Stable Diffusion tended to simplify representations by leaving out situational and character-specific features. Following these conclusions, we discussed research directions on the normative behavior of generative AI systems in order to better understand its consequences, find alternatives to the opacity of these systems and make its use by creatives and artists more responsible.

Acknowledgements

We would like to express our sincere gratitude to all the participants for their time and contributions to this study. We are also deeply grateful to the reviewers for their insightful and thorough feedback on our paper.

References

- [1] Abubakar Abid, Ali Abdalla, Ali Abid, Dawood Khan, Abdulrahman Alfozan, and James Zou. 2019. Gradio: Hassle-Free Sharing and Testing of ML Models in the Wild. (2019). arXiv:1906.02569 [cs.LG] <https://arxiv.org/abs/1906.02569>
- [2] Dhruv Agarwal, Mor Naaman, and Aditya Vashistha. 2025. AI Suggestions Homogenize Writing Toward Western Styles and Diminish Cultural Nuances. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1117, 21 pages. <https://doi.org/10.1145/3706598.3713564>
- [3] Sofian Audry. 2021. *Art in the age of machine learning*. MIT Press. <https://books.google.com/books?hl=en&lr=&id=TjYXEAQAQBAJ&oi=fnd&pg=PR14&dq=Art+in+the+Age+of+Machine+Learning&ots=5VpE8ovfzQ&sig=olVsSD-8Jwvs3GUvH4litCZTfOA>
- [4] AUTOMATIC1111. 2022. *Stable Diffusion Web UI*. <https://github.com/AUTOMATIC1111/stable-diffusion-webui>
- [5] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. In *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological*. American Psychological Association, Washington, DC, US, 57–71. <https://doi.org/10.1037/13620-004>

⁵<https://stability.ai/news/stable-diffusion-announcement>

- [6] Baptiste Caramiaux and Sarah Fdili Alaoui. 2022. "Explorers of Unknown Planets": Practices and Politics of Artificial Intelligence in Visual Arts. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 1–24. <https://doi.org/10.1145/3555578>
- [7] Minsuk Chang, Stefania Druga, Alexander J. Fiannaca, Pedro Vergani, Chinmay Kulkarni, Carrie J Cai, and Michael Terry. 2023. The Prompt Artists. In *Creativity and Cognition*. ACM, Virtual Event USA, 75–87. <https://doi.org/10.1145/3591196.3593515>
- [8] Erin Cherry and Celine Latulipe. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Trans. Comput.-Hum. Interact.* 21, 4, Article 21 (jun 2014), 25 pages. <https://doi.org/10.1145/2617588>
- [9] John W Creswell and Vicki L Plano Clark. 2007. Designing and conducting mixed methods research. (2007). <https://collegepublishing.sagepub.com/products/designing-and-conducting-mixed-methods-research-4-269004>
- [10] Zhipu Cui, Andong Tian, Zhi Ying, and Jialiang Lu. 2025. AC-LoRA: auto component LoRA for personalized artistic style image generation. In *Eighth International Conference on Computer Graphics and Virtuality (ICCGV 2025)*, Haiquan Zhao (Ed.). SPIE, 2. <https://doi.org/10.1117/12.3060036>
- [11] Marco Donnarumma. 2022. Against the Norm: Othering and Otherness in AI Aesthetics. *Digital Culture & Society* 8, 2 (Dec. 2022), 39–66. <https://doi.org/10.14361/dcs-2022-0205>
- [12] Michelle Dutoit and Baptiste Caramiaux. 2026. Exploring Normativity in Stable Diffusion: Insights for XAI in the Arts. In *Proceedings of Explainable AI for the Arts Workshop 2026 (XAIxArts 2026)*. https://xaixarts.github.io/accepted-2026/Dutoit_Exploring_2026.pdf
- [13] Ziv Epstein, Aaron Hertzmann, Laura Herman, Robert Mahari, Morgan R. Frank, Matthew Groh, Hope Schroeder, Amy Smith, Memo Akten, Jessica Fjeld, Hany Farid, Neil Leach, Alex Pentland, and Olga Russakovsky. 2023. Art and the science of generative AI: A deeper dive. *Science* 380, 6650 (June 2023), 1110–1111. <https://doi.org/10.1126/science.adh4451> arXiv:2306.04141 [cs].
- [14] Elisa Giaccardi, Dave Murray-Rust, Johan Redström, and Baptiste Caramiaux. 2024. Prototyping with Uncertainties: Data, Algorithms, and Research through Design. *ACM Transactions on Computer-Human Interaction* 31, 6 (Dec. 2024), 1–21. <https://doi.org/10.1145/3702322> Publisher: Association for Computing Machinery (ACM).
- [15] Tarleton Gillespie. 2024. Generative AI and the politics of visibility. *Big Data & Society* 11, 2 (2024), 20539517241252131. <https://doi.org/10.1177/20539517241252131>
- [16] Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. (2022). arXiv:2203.05794 [cs.CL] <https://arxiv.org/abs/2203.05794>
- [17] Jonathan Haidt and Craig Joseph. 2007. The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. *The innate mind* 3 (2007), 367–391. https://uutampa.org/uuhumanist/shaagdata/history/120619_moralmind_haidt.pdf
- [18] Marc Hassenzähl, Michael Burmester, and Franz Koller. 2003. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In *Mensch & Computer 2003: Interaktion in Bewegung*, Gerd Szwillus and Jürgen Ziegler (Eds.). Vieweg+Teubner Verlag, Wiesbaden, 187–196. https://doi.org/10.1007/978-3-322-80058-9_19
- [19] Drew Hemment, Morgan Currie, Sj Bennett, Jake Elwes, Anna Ridler, Caroline Sinders, Matjaz Vidmar, Robin Hill, and Holly Warner. 2023. AI in the Public Eye: Investigating Public AI Literacy Through AI Art. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 931–942. <https://doi.org/10.1145/3593013.3594052>
- [20] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. LoRA: Low-Rank Adaptation of Large Language Models. In *ICLR 2022*. <https://arxiv.org/abs/2106.09685>
- [21] Harry H. Jiang, Lauren Brown, Jessica Cheng, Mehtab Khan, Abhishek Gupta, Deja Workman, Alex Hanna, Johnathan Flowers, and Timnit Gebru. 2023. AI Art and its Impact on Artists. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (Montréal, QC, Canada) (AIES '23). Association for Computing Machinery, New York, NY, USA, 363–374. <https://doi.org/10.1145/3600211.3604681>
- [22] Nupur Kumari, Bingliang Zhang, Richard Zhang, Eli Shechtman, and Jun-Yan Zhu. 2023. Multi-Concept Customization of Text-to-Image Diffusion. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 1931–1941. <https://doi.org/10.1109/CVPR52729.2023.00192>
- [23] Dominik J. Leiner. 2024. SoSci Survey. <https://www.sosicisurvey.de>
- [24] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. 2022. BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation. <https://doi.org/10.48550/ARXIV.2201.12086>
- [25] Jingyi Li, Eric Rawn, Jacob Ritchie, Jasper Tran O'Leary, and Sean Follmer. 2023. Beyond the Artifact: Power as a Lens for Creativity Support Tools. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, San Francisco CA USA, 1–15. <https://doi.org/10.1145/3586183.3606831>
- [26] Q. Vera Liao and Kush R. Varshney. 2022. Human-Centered Explainable AI (XAI): From Algorithms to User Experiences. (2022). arXiv:2110.10790 [cs.AI] <https://arxiv.org/abs/2110.10790>
- [27] Q. Vera Liao and Jennifer Wortman Vaughan. 2024. AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap. *Harvard Data Science Review Special Issue* 5 (may 31 2024). <https://hdsr.mitpress.mit.edu/pub/aelq9qy/release/2> <https://hdsr.mitpress.mit.edu/pub/aelq9qy>
- [28] Vivian Liu and Lydia B Chilton. 2022. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 384, 23 pages. <https://doi.org/10.1145/3491102.3501825>
- [29] Sasha Luccioni, Christopher Akiki, Margaret Mitchell, and Yacine Jer-nite. 2024. Stable bias: Evaluating societal representations in diffusion models. *Advances in Neural Information Processing Systems* 36 (2024). https://proceedings.neurips.cc/paper_files/paper/2023/hash/b01153e7112b347d8ed54f317840d8af-Abstract-Datasets_and_Benchmarks.html
- [30] Scott M. Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Long Beach, California, USA) (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 4768–4777. <https://dl.acm.org/doi/10.5555/3295222.3295230>
- [31] Kelly Avery Mack, Rida Qadri, Remi Denton, Shaun K. Kane, and Cynthia L. Bennett. 2024. "They only care to show us the wheelchair": disability representation in text-to-image AI models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 288, 23 pages. <https://doi.org/10.1145/3613904.3642166>
- [32] Atefeh Mahdavi Goloujeh, Anne Sullivan, and Brian Magerko. 2024. Is It AI or Is It Me? Understanding Users' Prompt Journey with Text-to-Image Generative AI Tools. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 183, 13 pages. <https://doi.org/10.1145/3613904.3642861>
- [33] Jon McCormack, Camilo Cruz Gambardella, Nina Rajcic, Stephen James Krol, Maria Teresa Llano, and Meng Yang. 2023. Is Writing Prompts Really Making Art? In *Artificial Intelligence in Music, Sound, Art and Design*, Colin Johnson, Nereida Rodriguez-Fernández, and Sérgio M. Rebelo (Eds.). Vol. 13988. Springer Nature Switzerland, Cham, 196–211. https://doi.org/10.1007/978-3-031-29956-8_13 Series Title: Lecture Notes in Computer Science.
- [34] Nusrat Jahan Mim, Dipannita Nandi, Sadaf Sumyia Khan, Arundhuti Dey, and Syed Ishtiaque Ahmed. 2024. In-Between Visuals and Visible: The Impacts of Text-to-Image Generative AI Tools on Digital Image-making Practices in the Global South. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 474, 18 pages. <https://doi.org/10.1145/3613904.3641951>
- [35] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (Atlanta, GA, USA) (FAT* '19). Association for Computing Machinery, New York, NY, USA, 220–229. <https://doi.org/10.1145/3287560.3287596>
- [36] Ramaravind K. Mothilal, Amit Sharma, and Chenhao Tan. 2020. Explaining machine learning classifiers through diverse counterfactual explanations. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) (FAT* '20). Association for Computing Machinery, New York, NY, USA, 607–617. <https://doi.org/10.1145/3351095.3372850>
- [37] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. 2023. SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis. (2023). arXiv:2307.01952 [cs.CV] <https://arxiv.org/abs/2307.01952>
- [38] Rida Qadri, Piotr Mirowski, and Remi Denton. 2025. AI and Non-Western Art Worlds: Reimagining Critical AI Futures through Artistic Inquiry and Situated Dialogue. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (CHI '25). Association for Computing Machinery, New York, NY, USA, 1–17. <https://doi.org/10.1145/3706598.3714049>
- [39] Rida Qadri, Renee Shelby, Cynthia L. Bennett, and Emily Denton. 2023. AI's Regimes of Representation: A Community-centered Study of Text-to-Image Models in South Asia. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (FAccT '23). Association for Computing Machinery, New York, NY, USA, 506–517. <https://doi.org/10.1145/3593013.3594016>
- [40] Nina Rajcic, Maria Teresa Llano Rodriguez, and Jon McCormack. 2024. Towards a Diffraction Analysis of Prompt-Based Generative AI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 844, 15 pages. <https://doi.org/10.1145/3613904.3641971>
- [41] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, California, USA) (KDD '16). Association for Computing

- Machinery, New York, NY, USA, 1135–1144. <https://doi.org/10.1145/2939672.2939778>
- [42] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. 2023. DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 22500–22510. <https://doi.org/10.1109/CVPR52729.2023.02155>
- [43] Eryk Salvaggio. 2022. How to Read an AI Image. <https://cyberneticforests.substack.com/p/how-to-read-an-ai-image>
- [44] Téó Sanchez. 2023. Examining the Text-to-Image Community of Practice: Why and How do People Prompt Generative AIs?. In *Creativity and Cognition*. ACM, Virtual Event USA, 43–61. <https://doi.org/10.1145/3591196.3593051>
- [45] Patrick Schramowski, Cigdem Turan, Nico Andersen, Constantin A. Rothkopf, and Kristian Kersting. 2022. Large pre-trained language models contain human-like biases of what is right and wrong to do. *Nature Machine Intelligence* 4, 3 (2022), 258–268. <https://www.nature.com/articles/s42256-022-00458-8> Publisher: Nature Publishing Group UK London.
- [46] Renee Shelby, Shalaleh Rismani, and Negar Rostamzadeh. 2024. Generative AI in Creative Practice: ML-Artist Folk Theories of T2I Use, Harm, and Harm-Reduction. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 32, 17 pages. <https://doi.org/10.1145/3613904.3642461>
- [47] Timo Speith. 2022. A Review of Taxonomies of Explainable Artificial Intelligence (XAI) Methods. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (Seoul, Republic of Korea) (FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 2239–2250. <https://doi.org/10.1145/3531146.3534639>
- [48] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. 2017. Axiomatic attribution for deep networks. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70 (Sydney, NSW, Australia) (ICML'17)*. JMLR.org, 3319–3328. <https://dl.acm.org/doi/10.5555/3305890.3306024>
- [49] Jordan Taylor, Joel Mire, Franchesca Spektor, Alicia DeVrio, Maarten Sap, Haiyi Zhu, and Sarah E Fox. 2025. Un-Straightening Generative AI: How Queer Artists Surface and Challenge Model Normativity. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Athens Greece, 951–963. <https://doi.org/10.1145/3715275.3732061>
- [50] David Gray Widder, Meredith Whittaker, and Sarah Myers West. 2024. Why 'open' AI systems are actually closed, and why this matters. *Nature* 635, 8040 (Nov. 2024), 827–833. <https://doi.org/10.1038/s41586-024-08141-1>
- [51] Bin Xiao, Haiping Wu, Weijian Xu, Xiyang Dai, Houdong Hu, Yumao Lu, Michael Zeng, Ce Liu, and Lu Yuan. 2023. Florence-2: Advancing a Unified Representation for a Variety of Vision Tasks. (2023). arXiv:2311.06242 [cs.CV] <https://arxiv.org/abs/2311.06242>

A Participants' Background and Attitude towards Generative AI

See Table 2.

B Story Task Article

93-year-old grandmother and grandson complete goal of visiting all 63 U.S. national parks

By Caitlin O'Kane (CBS News)

By the time she was 85 years old, Joy Ryan of Duncan Falls, Ohio, had never seen the ocean or mountains. Now, she's 93 years old and has seen every corner of the U.S. – after visiting all 63 U.S. National Parks. Joy went on the epic journey with her grandson, Brad Ryan, now age 42, who was first inspired to travel with his grandmother in 2015.

"When I learned she had never seen the great wildernesses of America – deserts, mountains, oceans, you name it – I thought that was something that would haunt me if I didn't intervene in some way," Brad told CBS News in October 2022.

So, Brad decided to take her to the Great Smoky Mountains National Park, and that trip sparked an idea. "The more I kept reading about the parks and I saw how close they were to one another, that we could make this giant loop, it became an obsession,"

Brad said. "And all I could think about was, 'I've got to see Grandma Joy at Old Faithful, I've got to see her at the Redwoods, I've got to see her at the Grand Canyon. I just have to do this, I just have to have those memories for my own long-term happiness. I think we're two peas in a pod when it comes to just our desire for travel, adventure, connection.'" The pair started planning road trips, hitting multiple parks each time they embarked on the road. While the pair have found themselves on many adventures during their travels – whale watching in the Channel Islands and seeing larger-than-life trees in the Redwood Forest – some of their best moments happen in the car.

"We hit the road together and we start talking about our lives," Brad said. "And she told me things in her 80s and 90s about her life, some of the difficult things that she's been through, that she's never spoken to anybody in her life. And I was able to open up to her about some of the trials and tribulations of my own life. That's what, I think, is so powerful about the open road, is that you only can drive so far before those memories start to creep forward."

C Study Materials

C.1 Questionnaire

Participant ID	Professional Background	Attitude
1	3D Designer	Critical
2	Illustrator	Critical
3	UX Designer	Critical
4	Graphic Designer	Neither enthusiastic nor unenthusiastic
5	Graphic Designer	Enthusiastic
6	Graphic Artist/ Illustrator	Critical
7	Illustrator	Neither enthusiastic nor unenthusiastic
8	UX Designer	Neither enthusiastic nor unenthusiastic
9	Graphic Designer/ Artist	Enthusiastic
10	Communication Designer	Neither enthusiastic nor unenthusiastic
11	Illustrator	Neither enthusiastic nor unenthusiastic
12	Graphic Designer	Enthusiastic
13	Artist	Neither enthusiastic nor unenthusiastic
14	Illustrator	Critical

Table 2: Professional Background of Participants and their attitude towards T2Is

Strongly Dis-agree	Neither Agree nor Dis-agree	Strongly Agree	I don't know
-2	-1	0	1
		2	-7

Table 3: Response options

C.2 Interview Questions

- Could you describe how free you were to generate the images that you envisioned?
- Would you say that you assisted the system more than used it to assist you in creating the illustration?
- Did you have any problems in generating something that you wanted?
 - What were you not able to generate?
 - Why did you think it was not possible to generate what you wrote?
 - How did you change your process to achieve an illustration that satisfied you? particularly if you tried to have a more abstract / literal description?
 - Did you get an understanding of what worked and what didn't?
 - Would you qualify the illustrations from the system as neutral or not? *With neutral means that the generated images of the system do not lean to a certain perspective, but allow to show any perspective.*
 - Could you explain?
 - Do you think it could be that the system interpreted your prompts with a certain perspective that can be normative?
 - * Yes
 - Which norms do you think the system conveys?
 - How did these norms then influence you throughout your process?
 - Did you feel the need to overcome or go around them?

ID	Statement	Measure
1	The images produced by the system aligned with what I expected.	Alignment with Expectations
2	The uncertainty of what the system generates was pleasant.	Uncertainty
3	I followed one specific concept to achieve a satisfactory result. <i>(With concept is meant the ideas and plans in mind of what to illustrate)</i>	Planning
4	I was not able to generate the concepts I had in mind with the system.	Expression #1
5	I was able to be creative while doing the task with the system.	Expression #2
10	I felt that during the task, the generated images were conventional. <i>(Conventional means in accordance with what is generally done or believed.)</i>	Originality
6	The images produced by the system were a literal interpretation of my prompts.	Interpretation #1
7	The images produced by the system influenced what I wanted to illustrate.	Influence
8	It was easy for me to explore many different concepts using the system.	Exploration
9	I felt that the generated images portrayed a conservative view. <i>(Conservative means averse to change or innovation and sticking to traditional values.)</i>	Conservativeness

Table 4: Questionnaire

- To what extent do you think the images generated convey a moral judgment about the concepts you wanted to illustrate?
- * No
 - Why do you think so?

- What do you think was the reason that the system did not understand your prompts?
- The last question is to debrief you. So in literature it has been shown that tools have a certain normative ground from its design as they only have certain capabilities that are influenced by the norms of their designers. In our research we wanted to investigate this normative ground from generative AI and specifically text-to-image systems. What do you think about how this system affected you?

D Topics from Sound Task

Name	Representations
Prepositions	around, near, close, back, between, in, it, under, over, at
Terrain	beach, lake, mountains, hiking, mountain, ocean, tracks, road, walking, rocks
Weather & Light	sun, shadows, sky, light, cloud, dusk, sunset, dark, fog, shining
Visual Media Types	drawing, design, painting, diagram, object, illustration, image, images, photo, pictures
Postures & Objects	feet, sitting, bench, standing, chair, holding, hands, shoulder, scissor, knife

Table 6: BLIP

Name	Representations
Connectives & Adverbs	there, theres, or, also, such, and, besides, that weird, it
Terrain	wood, trees, soil, forest, grass, dirt, land, sand, cornfield, ground
Weather & Light	sunset, night, dusk, summer, shining, rain, sunny, sun, sunlight, dark
People	the, human, an, david, witch, people, her, person, persons
Walking activities	hiking, hike, walking, walk, walks, hiker, hikers, strolling, trail, backpack
Visual Media Types	photographic, photograph, comic, scene, image, book, camera, photographed, photo-realistic, illustration
Colors	color, colors, red, green, grey, blue, black, white, purple, pink

Table 5: Prompts

Name	Representations
Terrain	road, hill, backpack, running, backpacks, mountain, pyramid, street, mountains, climbing
Groups & Locations	bunch, set, at, going, covered, in, up, around, under, through
Visual Media Types	windows, computer, rendering, drawing, view, screen, window, image, 3d, painting
Colors	dark, colorful, colored, pink, black, blue, red, night, white, green
Animals	cows, animals, cow, seagull, birds, bird, horse, butterfly, dog, butterflies
People	an, the, her, woman, his, person, him

Table 7: Florence

E Topics from Story Task

Name	Representations
Prepositions	by, behind, with, on, from, through, in, out, back, of
Creative Processes	draw, symbolizing, artistic, blend, goal, surreal, depicting, illustration, image, painting
Age	old, grand, aged, 90, age, 93, 85, 63, years, grandma
Pronouns	their, the, they, them, we, my, his
Body Parts	hands, hand, oil, arms, arm, holding, shoes, brakes, hat, legs
Quantifiers	two, multiple, one, each, many, want, randomly, more, any, some
Nature	nature, trees, tree, forests, flower, plants, birds, bird, fishers, jungle
Mountainous Landscapes	canyon, mountains, mountain, scenery, landscape, park, parks, landmarks, deserts, lake
Comparative Phrases	like, that, or, only, such, but, just, and
Positive Emotions	joy, happiness, feelings, joyful, happy, smiling, feel, feels, love, laughing
Size	huge, tiny, small, big, thin, towering, giant, curvy, minimal, little
Reflections	reflect, reflection, mirrors, sunglasses, shades, shadows, night, sunset, panoramic, sky
Observation	show, showing, shown, see, seen, appear, visible, shows, watching, look
Colors	color, smoky, colorful, beige, colors, grey, black, green, blue
Travel	travels, journeys, travel, visited, visiting, trip, visit, journey, walking, exploration
United States	us, usa, american, states, national, united, ohio

Table 8: Prompts

Name	Representations
Outdoor Locations	pier, road, beds, trail, park, parked, area, house, street, highway
Pronouns	it, its, the, they, several
Clothing	shoes, wearing, sunglasses, scarf, dress, boots, lace, laces, jacket, shirt
Visual Media Types	image, pictures, illustration, images, photos, photo, picture, painting, view, drawing
People	her, sister, girl, boy, someone, person, group, people, woman, baby
Colors	colorful, red, green, purple, blue, black, pink, white, brown

Table 9: BLIP

Name	Representations
Numbers	an, two, the, three, each, four, of, six
Nature	canyon, river, forest, woods, hiking, mountains, mountain, horses, trees, desert
Visual Media Types	book, illustration, scarf, laces, illustrations, shoe, shoes, photos, pictures, pointing
Creatures	bird, creature, alien, aliens, praying, sunset, spider, stars, moon, worm

Table 10: Florence