

Exploring Normativity in Stable Diffusion: Insights for XAI in the Arts

MICHELLE DUTOIT, LMU Munich, Germany and ISIR, Sorbonne Université, CNRS, France

BAPTISTE CARAMIAUX, ISIR, Sorbonne Université, CNRS, France

Generative text-to-image (T2I) systems are increasingly adopted in creative practice, yet their normative behaviors remain under-explored from the perspective of creative practitioners. In this workshop paper, we present a within-subject study with 14 creative practitioners using Stable Diffusion to create illustration from two tasks of differing specificity. We investigate whether and how practitioners perceive normative behavior in a T2I system and how it impacts their creative process depending on task specificity. Our findings show that participants perceived normative behavior through invariant patterns, stereotypical output, and unsolicited omission or addition of details. These experiences led to feelings of disempowerment and creative compromise. We discuss implications for XAIxArts, including prompt transparency and artist empowerment in creative contexts.

Additional Key Words and Phrases: Normativity, Text-to-Image, Generative AI, Creative Practitioners

ACM Reference Format:

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)*. ACM, New York, NY, USA, 6 pages.

1 A Brief Background and Motivation

T2I systems have been shown to misrepresent cultures and reinforce Western-centric norms [1, 9]. 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 [23]. Ethnographic work in Bangladesh further illustrates how such systems can limit creative ideation and yield distorted or inaccurate representations for local artists [21]. This relates to normativity of generative AI (genAI) systems, which have been shown to tend to replicate normative narratives, and rarely producing non-normative alternatives when not mentioned explicitly [10]. We believe that XAIxArts could gain from an understanding of normativity. First, because the normative basis of the genAI tools leads to outputs that may not be understandable or transparent to the people using it. Understanding normativity and finding solutions through XAIxArts could help reduce its negative impacts and alleviate its influence. Second, studying how people perceive and deal with normativity progresses XAIxArts interventions. There has already been research on strategies for circumventing normative behaviors that have been shown to include repeated prompting, finding workarounds, iteratively refining prompts to achieve their objectives, and testing multilingual prompts [22, 33]. Further studying these strategies can provide insight on designing XAIxArts strategies.

Previous studies in the normative behavior of AI-based systems tend to illustrate the values and culture conveyed by these systems, and how users try to circumvent this gaze, but less so the processes through which these normative behaviors operate. This gap is particularly significant in creative and artistic contexts, where T2I systems are increasingly

Authors' Contact Information: Michelle Dutoit, michelle.dutoit@ifi.lmu.de, LMU Munich, Munich, Germany and ISIR, Sorbonne Université, CNRS, Paris, France; Baptiste Caramiaux, caramiaux@isir.upmc.fr, ISIR, Sorbonne Université, CNRS, Paris, France.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

adopted for ideation and production [25]. Artists develop strategies to achieve the desired results, such as creating prompt templates and adapting their language to match their expectations [7, 19]. Recent research also reported shortcomings in the use of T2I in creative tasks, including the difficulty in expressing thoughts in words [20], and “concept entanglement”, which refers to the system’s inability to represent specific concepts independently of other concepts [28].

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. Our research question is “**How do creative practitioners perceive the normativity of the T2I system depending on the task specificity, and how does the normative behavior affect them?**”.

2 Study Design

We recruited 14 visual creative practitioners (graphic designers, illustrators, UX designers with illustration background, and others) for a within-subject study located in Western Europe. The study was reviewed and approved by the institution’s Research Ethics Committee. Participants were recruited through the authors’ networks and calls advertised on social media. Our inclusion criteria were visual creative practitioners with no or minimal prior T2I experience, to avoid pre-formed strategies for circumventing T2I systems’ limitations and normative constraints.

The study was conducted in person or online using a video conferencing tool. In the beginning, the participants received an information letter and completed a consent form. Then participants received two tasks, each requiring them to create an illustration from a given source medium using the Stable Diffusion WebUI [2]. We chose to use Stable Diffusion, as local deployment gave us control over generation speed and removed prompt-usage restrictions. In the first task, the stimulus was a 30-second audio clip, evoking an outdoor hiking environment, featuring sounds of footsteps, birds, and wind. In the second task, participants received a short news article describing a man in his 40s and his grandmother in her 90s visiting all U.S. national parks together, which is disclosed in the Appendix A. We chose this story because it depicts a grandmother-grandson relationship with both of older age, a pairing that is rarely represented in illustrations and, therefore, difficult to generate with genAI. Tasks were presented in a fixed order rather than counterbalanced to prevent the narrative details of the story from influencing participants’ interpretation of the soundscape. This fixed order preserved exploratory freedom in the first task, though it may have introduced risks of order, practice, and fatigue effects. Finally, each session concluded with a semi-structured interview, which addressed how the participants perceived the normativity of the system, specifically by asking how they evaluated expressive freedom, assistance provided by the system, and perceived neutrality of the tool.

We analyzed the interview data as follows: The recordings were transcribed and translated locally on a secure computer with Whisper from OpenAI [24], as the interviews were conducted in English, German, and French. The interviews were edited for the paper in intelligent verbatim for readability and the translations were reread to ensure validity. Data were analyzed using inductive thematic analysis, as described in Braun and Clarke [3–5]. Both researchers familiarized themselves with the data, and independently created codes. They consolidated them together in a shared codebook, derived themes and sub-themes, and refined or eliminated them until a consensus was reached.

3 Highlights from our Analysis

Our findings indicate that participants identified normative behavior in the system through three main patterns: stereotypical imagery, invariant responses to prompts, and the omission or unsolicited addition of details.

First, participants identified the system’s normative behavior as coming from stereotyped, biased imagery. One participant for example observed that the system “*conveys the western culture norms, middle class*” and has “*a commercial, low key visual culture*.” This aligns with prior work documenting biases in generative models, including gender-profession associations [16] and Western-centric depictions of non-Western cultures [23]. Some example images are shown in Figure 1.

Participants also noticed repeated compositional patterns as another manifestation of the system’s normative behavior. Several participants observed shared commonalities in layout, perspective, and framing, although they often struggled to articulate them precisely. One participant probed these invariants deliberately, repeating the same prompt words and inspecting what remained consistent across outputs. Another way in which participants perceived normativity in the system’s behavior was through details specified in the prompts, but ignored by the model. Participants noted this in the interviews, describing normativity through omitted details, particularly in the more specific task (Story Task). For example, the male character was frequently left out or depicted as the same age as the older woman, as well as ignored style cues such as layout, line trace, and color. They theorized that the system performed better with less specified prompts, as too many details in prompts were often ignored. However, participants observed that when fewer details were included, the system generated imagery that conformed more closely to normative representations, such as that the older woman was generated with white hair, glasses, and a cane. One participant turned this into a deliberate strategy, abandoning the grandmother-grandson concept entirely to probe what non-normative content the system could produce.

Beyond perceiving the norms, it impacted the behavior of the participants. It contributed to a sense of disempowerment, as captured by one participant: “*So you could say that I assisted the tool, but I don’t know if the tool needed it.*”. A common response was to abandon efforts to influence aspects perceived as uncontrollable, raising the possibility that users may forgo non-normative creative goals when faced with the system’s constraints. Participants resorted to various compromises on their initial concept and tried to find a satisfactory outcome within the norms. They adjusted their concepts to match perceived system capabilities, adopted conventional prompt language (e.g. prompting a family portrait), or abandoned their initial concept altogether in favor of iterating on whatever the system first produced.



Fig. 1. Examples of Generated Images

4 Discussion

This work aims to contribute to the development of XAIxArts in three ways. First, we highlight XAI as an opportunity to go beyond default behavior. Second, we present prompt transparency as a promising research direction in this field. Third, we discuss the feeling of disempowerment stemming from the normative behavior of T2I and the extent to which XAI could prevent it.

Participants perceived normative behavior, which contributed to a simplification of visual representation that favored default depictions. This is significant as users tend to adhere to default options, representing limited perspectives and restricted visibility of alternative viewpoints [10], while artists use default abstractions and ‘make do’ with them, unable

to explore other possibilities [13]. Advancing research on default behaviors is therefore essential, including exploring strategies for moving beyond them, through e.g., leveraging rather than averaging out stochasticity to circumvent inherent biases [22], and by enabling deeper interaction with training data through techniques such as LoRA [11], or DreamBooth [27]. XAI can offer strategies to overcome default behaviors, such as interaction focused approaches [14] or black-box explanation tools [17, 26]. By allowing visibility and designing for glass-box abstractions [31], users can gather more insight into the nature of diffusion and its capabilities as their experience grows [30].

The system’s deviations from participants’ prompts, such as selectively removing or adding details without explanations, made participants reflect on the decisions made by the system and highlights the importance of prompt transparency. Such transparency aims to enable relevant stakeholders to develop an accurate understanding of the system’s capabilities, limitations, and mechanisms for controlling its outputs [15]. While transparency has primarily been investigated in high-stakes application domains, we argue that it deserves greater attention in the creative domain, given the role creative practice plays in shaping our cultural fabric. In a recent study, artists also discussed the benefit of transparency in fostering creative practice while informing about the appropriate uses of T2I [29]. Relevant to prompt transparency, existing techniques such as LIME [26], SHAP [17], and neural integrated gradients [32] 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. Researching prompt transparency could thus enable more effective application and appropriation of T2I models as creative tools and cultural instruments.

Participants in our study frequently expressed frustration and a sense of powerlessness when using the system. Mahdavi Goloujeh et al. [19] similarly found that users with focused objectives and well-defined visions often experienced frustration when the system failed to generate important elements of their visions, linking this to a loss of user autonomy. This is an interesting contrast, as genAI systems are typically marketed, for example, Stability AI describes Stable Diffusion as “empowering billions of people to create stunning art within seconds.”¹ Disempowerment through genAI remains underexplored. Previous research has documented how T2I output can disempower communities by perpetuating negative stereotypes and narratives of marginalized groups [18, 23], and more generally, the feeling of disempowerment and loss of autonomy can produce a *chilling effect* [12], and dissuade creatives from engaging with the development of creative-centered AI tools. The “democratization of art”¹ is negated through the harms these technologies bring to the creative community. Recent advances 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 necessary to identify effective adjustments [8]. 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. Li et al. [13] 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. This can be done by furthering research on the disempowerment, which directly relates to XAIxArts, as a central theme is the empowerment of creatives. Working directly with artists helps to design interventions that can specifically address them and their empowerment, such as developing their AI literacy, designing accessible and inclusive XAI tools, and engaging them for input and feedback [6].

¹<https://stability.ai/news/stable-diffusion-announcement>

References

- [1] 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*. 1–21.
- [2] AUTOMATIC1111. 2022. *Stable Diffusion Web UI*. <https://github.com/AUTOMATIC1111/stable-diffusion-webui>
- [3] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis*. American Psychological Association.
- [4] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11, 4 (2019), 589–597.
- [5] Virginia Braun and Victoria Clarke. 2021. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative research in psychology* 18, 3 (2021), 328–352.
- [6] Nick Bryan-Kinns, Shuoyang Zheng, Francisco Castro, Makayla Lewis, Jia-Rey Chang, Gabriel Vigliensoni, Terence Broad, Michael Paul Clemens, and Elizabeth Wilson. 2025. XAIxArts manifesto: explainable AI for the arts. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–8.
- [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. doi:10.1145/3591196.3593515
- [8] 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. doi:10.1117/12.3060036
- [9] Marco Donnarumma. 2022. Against the Norm: Othering and Otherness in AI Aesthetics. *Digital Culture & Society* 8, 2 (Dec. 2022), 39–66. doi:10.14361/dcs-2022-0205
- [10] Tarleton Gillespie. 2024. Generative AI and the politics of visibility. *Big Data & Society* 11, 2 (2024), 20539517241252131.
- [11] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. 2022. Lora: Low-rank adaptation of large language models. *ICLR* 1, 2 (2022), 3.
- [12] 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*. 363–374.
- [13] 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. doi:10.1145/3586183.3606831
- [14] Q Vera Liao and Kush R Varshney. 2021. Human-centered explainable ai (xai): From algorithms to user experiences. *arXiv preprint arXiv:2110.10790* (2021).
- [15] Q Vera Liao and Jennifer Wortman Vaughan. 2023. Ai transparency in the age of llms: A human-centered research roadmap. *arXiv preprint arXiv:2306.01941* (2023).
- [16] Sasha Luccioni, Christopher Akiki, Margaret Mitchell, and Yacine Jernite. 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
- [17] Scott Lundberg. 2017. A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874* (2017).
- [18] 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. doi:10.1145/3613904.3642166
- [19] 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. doi:10.1145/3613904.3642861
- [20] 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 Rodríguez-Fernández, and Sérgio M. Rebelo (Eds.). Vol. 13988. Springer Nature Switzerland, Cham, 196–211. doi:10.1007/978-3-031-29956-8_13 Series Title: Lecture Notes in Computer Science.
- [21] 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*. 1–18.
- [22] 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. doi:10.1145/3706598.3714049
- [23] 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. doi:10.1145/3593013.3594016
- [24] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. In *International conference on machine learning*. PMLR, 28492–28518.

- [25] Nina Rajcic, Maria Teresa Llano Rodriguez, and Jon McCormack. 2024. Towards a Diffractive Analysis of Prompt-Based Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–15.
- [26] 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*. 1135–1144.
- [27] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 22500–22510.
- [28] Téo 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. doi:10.1145/3591196.3593051
- [29] 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 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [30] Ben Shneiderman. 2007. Creativity support tools: accelerating discovery and innovation. *Commun. ACM* 50, 12 (2007), 20–32.
- [31] James Smith, Shm Garanganao Almeda, Timothy J Aveni, Anya Agarwal, and Bjoern Hartmann. 2026. Noise Pilot: Enabling Artistic Workflow Composition with Diffusion-Based Image Generation. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*. 1–20.
- [32] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. 2017. Axiomatic attribution for deep networks. In *International conference on machine learning*. PMLR, 3319–3328.
- [33] 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. doi:10.1145/3715275.3732061

A 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 is 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 here 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."