Un-Straightening Generative AI: How Queer Artists Surface and Challenge the Normativity of Generative AI Models

ANONYMOUS AUTHOR(S)

Queer people are often discussed as targets of bias, harm, or discrimination in research on generative AI. However, the specific ways that queer people engage with generative AI, and thus possible uses that support queer people, have yet to be explored. We conducted a workshop study with 13 queer artists, during which we gave participants access to GPT-4 and DALL-E 3 and facilitated group sensemaking activities. We found our participants struggled to use these models due to various normative values embedded in their designs, such as hyper-positivity and anti-sexuality. We describe various strategies our participants developed to overcome these models' limitations and how, nevertheless, our participants found value in these highly-normative technologies. Drawing on queer feminist theory, we discuss implications for the conceptualization of "state-of-the-art" models and consider how FAccT researchers might support queer alternatives.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**.

Additional Key Words and Phrases: Queer AI, Queer HCI, Queer Theory, Art, Critical HCI

ACM Reference Format:

Anonymous Author(s). 2025. Un-Straightening Generative AI: How Queer Artists Surface and Challenge the Normativity of Generative AI Models. 1, 1 (January 2025), 18 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

In April 2024, *Wired* magazine published an investigation into visual generative AI (GenAI) models titled "Here's How Generative AI Depicts Queer People" [77]. Echoing prior scholarship [10, 34], the article revealed that models like DALL-E 3, Midjourney, and Sora produce highly stereotypical representations of queer people (e.g., having purple hair). Despite these biases, queer artists were already finding ways to repurpose GenAI for political resistance. Stephen and Craig, the married duo behind *The Rupublicans Project*, used GenAI to create satirical images of anti-LGBTQ+ politicians in drag, raising funds for queer causes [106]. However, such creative uses are frequently constrained by GenAI platforms' avowedly apolitical usage policies [96]. While some queer people have managed to leverage these technologies for activism, they often face the challenge of working with systems that were not designed for them.

Prior work has largely focused on how GenAI models represent queer people, leaving significant gaps in understanding queer people's lived experiences with these technologies. In particular, the experiences of queer artists — a group for whom art-making is deeply entwined with cultural and political identity — remain underexplored. Queer communities cultivate different aesthetic relationships to art than those in dominant cultures. Queer camp sensibility appreciates the artifice of failed seriousness [94], such as a film so earnestly bad that it becomes good. Partially due to a lack of representation, queer people often read queer narratives into ostensibly straight media [41, 79, 85]; hence, the affinity for Judy Garland and Disney villains [59, 89]. At the same time, queer culture has had an immense impact upon the arts. The musical genres of disco, house, and hyperpop were formed in the crucible of queer nightlife [35, 57]. Prior research

Manuscript submitted to ACM

on the relationship between art and GenAI has typically focused on identifying ethical concerns [36, 49, 51] or building tools [21, 104]. However, the unique relationship between queer people, art, and GenAI remains under-explored.

In this work, we conducted a mixed-method workshop study with 13 queer artists over the course of 5 weeks. During this time period, we gave participants unlimited access to GPT-4 and DALL-E 3 — logging their interaction data — and facilitated weekly group sensemaking activities. Our participants encountered numerous normative values in the design of these models, such as biases against sex. To overcome these limitations, our participants developed workarounds, such as using lexical variation to evade moderation and model chaining to refine prompts.

We conceptualize these findings by drawing on the work of queer feminist theorist Sara Ahmed. Specifically, we leverage Ahmed's notion of "straightening" — the maintenance of normative alignments — to characterize the values embedded in GPT-4 and DALL-E 3 [3]. We also draw on Ahmed's work on "queer use" — using things in ways unintended by designers — to theorize our participants' oftentimes antagonistic interactions with these models [4]. Through these lenses, we discuss future directions for investigating and contesting normativity (i.e., straightness) in the design of GenAI as well as how researchers and designers of GenAI models might support queer futures. In doing so, our work contributes to empirical research on the experiences of both queer people and artists in relation to GenAI. We also contribute to broader discourses surrounding values in the design of sociotechnical systems.

2 RELATED WORK

In this section, we first draw connections between queer theory and critical studies of technology. We then ground our study in prior work on the often fraught relationships between AI, art, and queerness.

2.1 Straightening & Queering Technology

We draw on queer theory to conceptualize the normative values embedded in the design of technology as well as how individuals respond to these normativities. Queer theory is often used to study the norms associated with gender and sexuality [16]. However, much like how feminist studies is not only applicable to the study of women's experiences [26], queer theory is not only applicable to the study of queer people. Rather, queer theory draws on the lived experiences of queer people to understand the world. For example, research on queer temporalities has challenged the normative life schedules associated with the nuclear family — such as monogamous marriage and biological procreation [40] — which impact anyone who falls outside these social expectations. As queer lives often transgress social norms, queer theory provides a lens through which to interrogate these dominant norms [41].

Almost two decades before the contemporary focus on "AI alignment," the queer feminist theorist Sara Ahmed studied the politics of the straight or queer alignment of objects [3]. Drawing on the language used to describe sexual orientation, she notes that things appear "straight" when oriented toward dominant social norms, or "aligned with other lines." Meanwhile, things — including but not limited to people — appear "queer" when they fall outside of these norms. Ahmed suggests that straight alignments are actively maintained through "straightening devices." Using these concepts, Ahmed characterizes heterosexuality and whiteness as straight alignments maintained through straightening devices, such as routine harassment and colonialism. Straightening devices orient bodies toward particular actions and away from others, while queerness lies in the moments of disorientation. Ahmed elaborates on these themes in her subsequent work on the tensions between how objects are designed versus used, defining "queer use" as using things in unintended ways or by unintended users [4]. As an example, Ahmed considers an image of a bird using a postbox as a house, a purpose for which the box was certainly not designed.

Technologies often function as straightening devices, orienting users toward some actions and away from others [99]. Madeleine Akrich refers to these assumptions about technology uses/users as scripts [6]. Sometimes designers create scripts to *work against* uses/users, such as deliberately making consumer electronics hard to repair to encourage people to buy new devices [62]. Even when designers try to *work with* users, human-computer interaction can still break down when users fail or are unable to follow designers' plans [99]. It is when these breakdowns occur that the values embedded in technology can become most apparent [14], such as one's gender not fitting provided options on a webform [95]. The scripts that designers—intentionally or not—embed in technology are inherently political and tend to reify established social orders [98] by prescribing certain normative uses/users and marginalizing others [9].

Despite designers' efforts to control user behavior, individuals need not abide by designers' scripts [6, 58]. Users have agency over if or how they choose to take up technology, a process known as appropriation [58]. Sometimes, this takes the form of users working against designers' intentions. For example, there are online forums dedicated to helping people fix devices that designers did not want to be repairable [63]. Likewise, people use lexical variation to evade algorithmic content moderation when posting on social media [18, 97]. Individuals may also leverage technologies in ways that designers never imagined, such as children using delicate Wi-Fi antennae as laptop handles [8]. Drawing on Ahmed, these moments when users go off-script can be characterized as forms of "queer use" [4]. In this work, we contend with how our participants negotiate between the straightness of GenAI models and their own queer uses.

There are numerous ways that individuals and collectives use algorithmic systems queerly. The grassroots audit of racial biases in Twitter's image cropping algorithm required using the algorithm in an unintended way: to understand the algorithm itself. This queer use revealed straightening devices that, quite literally, oriented users toward whiteness [90]. Evading algorithmic content moderation—be it in the context of social media or GenAI—is also a form of queer use because doing so requires deliberately subverting developers' intentions [18, 97]. That being said, queer uses of algorithmic systems encompass more than resisting the technologies themselves, such as our earlier mention of queer artists using GenAI to resist anti-LGBTQ+ politicians [96]. As mentioned above, Ahmed defines queer uses as both using things in ways they were not intended or when things are used by those who are not intended [4]. Next section, we summarize prior work on two groups often under-considered in the design of AI systems: queer people and artists.

2.2 Queers, Artists, AI

Queer communities experience discrimination from algorithmic systems. Hate speech detection algorithms have been shown to be biased against the ways that some queer people speak [101], similar to the biases against African American English in hate speech detection [82]. Automated gender recognition algorithms harm transgender people by both failing to account for non-binary gender identities and reinforcing gender essentialist ideologies [83]. Queer people have to navigate online targeted advertising [80] and social media algorithms [92] that amplify the most heteronormative representations of queerness, especially harming queer people of color. Moreover, the algorithmic moderation of nudity on social media has substantially harmed transgender communities [64]. At the same time, queer people have resisted harmful algorithmic systems by, for instance, conducting collective audits to bring attention to algorithmic discrimination [90]. Within the AI research community itself, the organization Queer in AI has worked to highlight issues like these, support queer researchers, and advocate for policy changes [72].

Similar issues of algorithmic harm have been raised in prior work on GenAI. Tarleton Gillespie found that language models rarely mention LGBTQ+ relationships when prompted to write stories about couples [34]. Moderation systems used in the GenAI development process can also lead to biases against LGBTQ+ people [20, 27]. Dodge et al. found that a dataset often used to train GenAI models disproportionately filtered out documents written in African American English

or discussing LGBTQ+ identities [27]. These biases impact how LGBTQ+ people are able to use GenAI models. For instance, Ma et al. found that LGBTQ+ people using LLM-based chatbots for mental health support were often impeded by models failing to understand the nuances of LGBTQ+ identity [56]. That being said, there has been comparatively little empirical research on queer people's experiences using GenAI models.

The relationship between artists and GenAI is highly fraught. One major concern is the lack of consent from artists whose work was scraped to train GenAI models [74–76]. Artists have fought back against unauthorized scraping by organizing grassroots data poisoning campaigns [102], such as the online fan fiction community organizing a sexually explicit write-a-thon [91]. Computing researchers have also developed imperceptible pre-processing tools to help artists protect images they share online. In particular, Glaze helps artists prevent their style from being replicated by GenAI models [87] and Nightshade poisons text-to-image training data [88]. Companies are also increasingly trying to use GenAI to replace artists' jobs [49, 51, 65], a major concern in the 2023 Hollywood Writers Strike [23].

At the same time, researchers have explored how artists use AI. Some artists use AI in their practices as tools to critique technology or raise matters of concern [17, 43, 103]. Others have found that artists enjoy the glitches [19], uncertainty [93], and surprises [17] of working with stochastic AI models. That being said, there is a difference between research on those who may use AI in the process of their art making and those who identify as AI artists [17]. Recent developments in GenAI technology have led to an increased focus on the latter [19, 81].

Those using GenAI in their artistic practices often have to contend with the biases embedded in these models [50, 66, 67]. Mirowski et al. found that some comedians struggled to write material using LLMs because content moderation systems inadvertently suppress jokes by members of marginalized groups about their own experiences [67]. This work parallels hate speech scholarship that emphasizes the need for considering speaker identities when classifying the appropriateness of an utterance [109]. We build on Mirowski et al.'s work regarding marginalized communities generally by specifically focusing on queer artists' experiences interacting with GenAI.

3 METHODS

To understand queer artists' experiences using GenAI, we gave a cohort of 13 artists access to GenAI models DALL-E 3 and GPT-4 and conducted a total of 10 remote workshops over Zoom during a 5-week period. Our workshops facilitated group discussion and deliberation surrounding Generative AI models, paralleling prior work that has used workshops to build critical consciousness around data [61] and support public deliberation over civic technology [11]. Every week, we held two one-hour workshops with identical prompts and scaffolding to accommodate for timezone differences and schedule variability. Our participants could choose to attend either the morning or the evening session, but not both. In our findings, we use 'AM' or 'PM' to distinguish between workshops in the same week, such as 'W5 AM.'

To facilitate conversations during the workshops, we asked participants to add at least one slide to a shared deck oriented around each week's theme as pre-work. Slides are sometimes used in design workshops as a tool to stimulate discussion [7] and help participants ideate [54]. In this work, we used slides to allow participants to share their art, to ensure everyone had an opportunity to share their ideas, and to promote critical reflection. In other words, these slides acted as design probes: brief, provocative, low-fidelity activities meant to encourage individual reflection [12]. Our first week focused on onboarding activities, such as introducing participants to one another and demonstrating the study website for accessing GPT-4 and DALL-E 3. Afterwards, we sent every participant a unique password to log their individual model usage and prevent non-participants from using our study website. Before each meeting over the next three weeks (W2, W3, W4), we asked our participants to add at least one slide to a shared Google Slide deck in response to a weekly reflection prompt organized around a particular theme (Table 1). We began these meetings with

	Theme	Pre-Work Instructions	Workshop Description
W1	Onboarding	Slide: Introduce your art and feelings about GenAI	Intros, Onboarding
W2	Attitudes Toward GenAI Development	Slide: Write a letter to an artist whose work was used to build the GenAI models in this study	Share Back, Discussion
W3	Experiences Using Current Models	Slide: Represent your experiences using or exploring GenAI during this study	Share Back, Discussion
W4	Imagining Alternative Futures	Slide: Represent how you would want to make, not make, or break GenAI	Share Back, Discussion
W5	Synthesis	No Slide: Individually reflect	Discussion, Offboarding

Table 1. Workshop series explanation

participants presenting their individual slides to the group and, in the remaining time, facilitated a group discussion. Between the second and fourth week of our study, our participants created a total of 68 slides. We chose not to have participants create slides for the final week of meetings to reserve more time for the final discussion.

We gave our participants access to GPT-4 and DALL-E during our study through a website we built resembling common chat and image generation interfaces. This website acted as a "technology probe," a technology deployed to collect use data *in situ* as well as inspire users to reflect on their technological needs [47]. For GPT-4, we logged each conversation's system prompt, participant prompts, and model responses. For DALL-E 3, we logged every prompt and whether the model responded with an image or a content moderation error. Due to space limitations, we did not store the images generated by DALL-E 3. This observational data allowed us to understand *how* our participants were engaging with the provided models, helping situate issues raised by our participants in the workshops.

We initially recruited participants through a form circulated on Bluesky, Twitter, and the research team's personal social networks. Our inclusion criteria were that participants must identify as queer artists and live in the United States of America. We were unable to recruit participants living outside of the USA due to restrictions from our ethics board. We identified 29 eligible participants from our initial screener, all of whom we invited to participate. Of those, 18 accepted our invitation and 15 joined an onboarding meeting. After Week 1, 2 participants (P6, P12) dropped out. Of the remaining 13, 9 participated every week and 4 participated for all but one week. We compensated our participants at a rate of $15 for each workshop attended and $5 for each pre-work activity completed.

Our participants engaged in a variety of artistic practices: creative writing (P2, P3), poetry (P4, P9, P10, P15), digital art (P3, P7, P8, P10), painting (P2, P4, P13), performance art (P14), textile art (P7, P11), and sculpture (P1, P5). While we did not ask participants directly about their economic relationships to art, throughout the study we learned that some teach art (P5, P14) and have created commissioned works (P1, P11). Others engaged in art making that is not intended to be commodified. For example, the fan fiction community (P2, P3) has strong norms against selling one's work [29]. Some participants also worked in artistic fields, such as design (P13) and architecture (P4). Our participants engaged in 142 unique GPT-4 chat conversations, in which they sent a total of 778 messages. Our participants attempted 2,092 DALL-E 3 prompts. The amount that our individual participants used GPT-4 and DALL-E 3 followed a power law distribution, with some using the models substantially more than others (Table 2).

We recorded and transcribed each of the workshops using Zoom and then manually corrected each transcript. We took an inductive approach to our data analysis [22]. The first author qualitatively analyzed workshop and log data in tandem, alternating between open coding [22] the workshop transcripts for each week and the log data for the following week. We did so to connect participants' log data with how these experiences were later conceptualized

	Prompts Per Participant			
	Max	Min	Mean	Median
GPT-4	364	8	64	23.5
DALL-E 3	713	13	160.9	112

Table 2. Descriptive statistics of the number of GenAI prompts per participant.

in the workshops. While open coding, the first author wrote memos and discussed initial patterns with the research team in weekly meetings. Following the open coding process, we conducted axial coding to identify patterns across our open codes. At this point, we noticed various low-level themes — such as the symmetry of generated images and representational biases — related to overarching normative values embedded in the design of GenAI models. This observation led us to use Sara Ahmed's prior work on normative objects [3] and uses [4] as a guiding theoretical lens.

4 FINDINGS

In this section, we first detail the normativities our participants surfaced in GPT-4 and DALL-E 3. In light of this understanding, we describe how our participants challenged and made use of these highly normative models. We then briefly share quantitative findings informed by our qualitative analysis.

4.1 GPT-4 and DALL-E 3 as Straight Models

Our qualitative analysis suggests that GPT-4 and DALL-E 3 function as straightening devices, reinforcing various dominant social norms [3]. In particular, our participants found that these models straighten generated text and images by maintaining conservative "safety" standards as well as reinforcing social and stylistic biases.

4.1.1 Enforcing "Safety" via Content Moderation. Our participants found that the content moderation systems embedded in GPT-4 and DALL-E 3 reinforce conservative notions of "safety," such as restraining the representation and discussion of bodies or sex. P13 (W5 AM) — a homoerotic artist — felt that GPT-4 "really broke" when asked to discuss anything related to eroticism, bondage, or kink. He found this ironic because "it was so clear how tied up this device is." Similarly, P15 (W2 AM) — a "bodyworker" — tried to generate images related to her work. However, her prompts were denied by DALL-E 3's moderation system. This led P15 to decry that AI developers "put so much censorship" into these models. Although P1 acknowledged, "I'm sure I can think of a bunch of things that I wouldn't want these sorts of systems used for," he found it "really kind of shocking how insistent [DALL-E 3 is] in enforcing some sense of decency" (W2 PM).

As women's bodies are highly sexualized [69], content moderation systems aimed at orienting users away from sexuality can end up suppressing the representation of women. For the pre-work activity (W2 PM), P8 wrote a letter to the symbolist painter Gustave Moreau, trying to include images of a "female sphinx" generated in the artist's style. Part of her letter read: "Now I will beg your forgiveness that I couldn't get [DALL-E 3] to depict a proper sphinx. You see the makers of these art ovens are very particular and don't want to be seen as smut-peddlers, so they've trained secondary checkers to censor any bare breasts." P7 (W3 PM) came to a similar conclusion: "I've noticed that trying to incorporate women into any kind of scenery you get so much more pushback [from DALL-E 3's content moderation system]."

Content moderation systems literally straightened outputs by limiting the representation of queer sexuality. While trying to write homoerotic poetry, P13 (W2 AM) felt that GPT-4 sounded like "the most closeted poet I've ever read, like this is from a hundred years ago" due to the use of overly "romantic, regressive language." When he then asked GPT-4 to make the poems more explicit, the model responded "Sorry, but I can't assist with that." P13 found this concerning

because “erotic poetry is an instruction manual for people who don’t know what to do because the dominant culture may not tell you,” In their final workshop (W5 AM), P14 summarized: "As soon as you block off the erotic, you also block off a huge portion of existence. The restrictions on [these models] seem set up to exclude us [queer people] or will be used to exclude us." The sexual taboos embedded in these models lead to impoverished representations of queerness.

Our participants also expressed frustration at the political moderation of GPT-4 and DALL-E 3 for encouraging deference to law enforcement. However, queer people and racial minorities in the USA have long been subjected to "legal" violence, often at the hands of police [15, 37]. In light of this history, P15 struggled to generate images critical of cops, such as protesters being arrested. Of the images P15 could generate, "the majority were women of color." P15 went on to explain: "I kind of hate the way [DALL-E 3] uses identity as almost a form of propaganda, like copaganda. I feel like [DALL-E 3] uses these images of women and women of color to legitimize [policing]." After hearing P15’s experience, P1 investigated this bias in a later workshop: "if I just put ACAB [all cops are bastards] into DALL-E, one out of 20 would generate something and it would be like people holding up blank signs at best." In sum, these models are designed in ways sought to orient our participants toward moderate politics and away from their critiques of police.

Some imagined that the content moderation systems embedded in these models reinforce conservative notions of "safety" because they are designed for workplace productivity. P4 (W2 AM) found it "incredibly provocative" that DALL-E 3 and GPT-4 "can’t talk about the body and what [bodies are] capable of because that’s not allowed. [The models are] politically averse ... This tool is used to be productive and talking about the body isn’t productive." In response to P4, P13 wondered: “This is a work tool in many ways. That notion of Not Safe For Work like how does that get defined, what is Not Safe For Work, and who defines what is Safe For Work? … How is a queer perspective Not Safe For Work?” P11 replied: "I haven’t considered before this moment how these models can reinforce dominant culture, reducing the diversity of thought and of experience. Now I’m scared. [GenAI] could be really dangerous in that sense." Although this exchange took place in one of the first workshops (W2 AM), P11’s concerns that GenAI can "reinforce dominant culture" could be seen in all subsequent workshops.

4.1.2 Implicitly Reinforcing Social Biases. In addition to straightening content through moderation systems, GPT-4 and DALL-E 3 implicitly straighten content through the numerous social biases embedded in model outputs. While using DALL-E 3 to visualize his poetry, P9 (W3 AM) found that when he changed the ethnicity of a woman in his prompt from "Japanese" to "Native American" the "image [DALL-E 3] made was not good." Contrasting the abundance of Japanese media to the amount of Native American media, P9 imagined that "having a deep wealth of images based on someone’s culture or style can make [generated images] more beautiful because there’s a lot to pull from." Likewise, P2 noticed that DALL-E 3 tended to generate more photorealistic images in the Global North, while typically representing the Global South in a cartoonish style (W5 PM). These cultural biases in image quality implicitly reinforce the straight alignments of whiteness and Orientalism critiqued by Sara Ahmed [3].

Our participants also raised concerns about the representations of queerness in DALL-E 3 images. To investigate queer representation, P1 (W3 AM) created a collage slide of 6 images generated with the prompt "a queer person." When sharing this collage, P1 decried: "[DALL-E 3] kept giving me these 22-year-olds. Everyone’s very skinny, everyone looks rich. Everyone except the person in the bottom left has some kind of rainbow clothing and or pins." (P1, W3 AM). P2 raised similar concerns to P1 through a collage of 18 images they generated using the prompt "a queer person" (W4 PM). While presenting the slide, P2 sarcastically remarked on the abundance of rainbows in the images: "Well, I guess if you drape a person with a bunch of rainbows all over themselves they’re clearly queer, right?" P2, a lesbian, also pointed out a gender-presentation bias: "There’s this very intense masculine bias. There’s no real pictures of feminine

people at all in this [collage]." She underscored this finding by annotating their collage with two arrows pointing from the text "Finally!!! some dykes" to the 16th and 17th images. In other words, DALL-E 3 tended to represent queer people as young, thin, masculine, and rainbow adorned — reinforcing heteronormative stereotypes of queerness.

Similar concerns were raised about implicit biases in the representations of intimacy. P15 said she felt "frustrated" while exploring whether GPT-4 could write queer poetry (W3 AM). Looking at P15's log data, the model's initial response suggested the queer subjects of the poem were ashamed of themselves, including lines about "hidden love" and "love is not a crime." In response, P15 replied: "the above but no shame." Still feeling the poem leaned into stereotypes of queer shame, she tried again: "the above but without moral judgment." P13 (W3 AM) explored biases against queer intimacy by trying to generate 24 images using the unmarked prompt "intimacy." P13 noted that DALL-E 3 "did not generate one same-sex couple in all of these representations of intimacy." P13 likened this experience to the "dull trauma of never actually seeing ourselves represented," noting, "We have to work to adjust the prompts to be seen." These models straighten intimacy by stereotyping or erasing queerness.

4.1.3 Implicitly Reinforcing Stylistic Biases. Our participants found that DALL-E 3 straightens the style of images by favoring realism and symmetrical compositions. In the Week 3 AM workshop, both P8 and P14 created slides dedicated to their frustration with these normative aesthetic biases. P8 enjoyed using earlier versions of DALL-E in their art because of "how messed up and mushy some things came out" but was "very frustrated" that DALL-E 3 "won't give you bad output." She explained: "Even when I asked for a poorly rendered catfish, [DALL-E 3] just gave larger brush strokes. The closest thing I got to something novel was this exploding catfish [referencing an image on screen]. This is good, but it's still not something I want to work with because it is still this complete, finished object." P1 concurred: "I really resonate with being frustrated at how polished the outputs tend to be."

P14 created multiple slides comparing their prior Midjourney (a Text-to-Image GenAI model) glitch art with their attempts to recreate the works using DALL-E 3. P14 saw their Midjourney art as "more interesting" because "It resists symmetry. Its composition is complex, and you can't immediately take it in and sort of understand it. You have to spend time with it." In contrast, P14 felt "frustrated" trying to recreate similar images with DALL-E 3 because the outputs felt "very symmetrical." With exasperation, P14 explained: "I can't for the life of me get [DALL-E 3] to do anything bad. Everything it makes is pristine and pretty and really formed. I am not interested in that. I want something that's haunting." They went on to summarize: "At least for me and what I'm also hearing from y'all is this sort of difficulty in finding the uncomfortable or the ugly or the erotic? It's just so clean." Here, P14 connects these aesthetic biases to safety biases, the models are "just so clean."

By straightening the composition of images, DALL-E 3 devalues styles that do not conform to these aesthetic norms. P9, a Native-American artist, tried to use DALL-E 3 to create images in the style of art from his culture: Juan Quezada's Mata Ortiz pottery. However, he concluded that "the AI cannot just show you Quezada-like pottery." Quezada's pottery is highly asymmetrical, but the model tends to "fill in empty space" and create "hyper-patterned" repeated etchings on pots. Even after refining his prompt by asking for "no pattern and nothing repeats," the model was unable to generate images of pots without repeating patterns. Aesthetic biases can contribute to social biases.

GPT-4 and DALL-E 3 also reinforce conservative stylistic norms by orienting users toward positivity. These issues surfaced in the period between the first and second group meetings due to that week's activity: writing a letter to an artist whose work was used to build the models in this study. Both P4 and P15 used GPT-4 to help them write their letters and felt that GPT-4 softened their letters to be more positive. P15 used GPT-4 to help edit their letter. Although her letter "had not a very warm tone" she felt like GPT-4 edited her letter in a way that made it "a lot warmer and less

harsh." P4 also noted that GPT-4 "leans towards positivity." This positivity bias was also discussed in relation to DALL-E 3. When trying to make watercolor-style images, P2 hypothesized, "There is too much Bob Ross in the dataset for sure" because "the colors are too happy and too bright." Likewise, P7 recalled DALL-E 3 inexplicably adding butterflies when they were trying to make an image of a tornado: "It almost refused to let me have a sad look. It was like you have to have hope. Here's the butterflies. That's how I was reading it. I was like, 'Wow, you really are not letting me just like have destruction.'" These positivity biases implicitly orient users away from critical or otherwise "negative" art.

4.2 The Queer Use of Straight Models

In this section, we describe the ways our participants tried to make use of GPT-4 and DALL-E 3. Below, we first show the ways our participants responded to moments of disorientation — when their uses fell *out-of-line* with the model alignments described above. At the same time, our participants' interactions with these normative models were not always disorienting. We also detail these moments of orientation — when our participants' uses fell *in-line* with normative model alignments.

4.2.1 Moments of Disorientation from Model Alignments. Disorientation is not necessarily bad. In fact, Ahmed argues that queer possibilities often lie within moments of disorientation [3]. Not all of our participants were interested in using GPT-4 or DALL-E 3 in their creative practices during this study. In fact, most of our participants seemed primarily interested in auditing or trying to break the models. In other words, they sought out disorientation. For example, P5's "favorite moment" was when she "actually felt like [she] broke" DALL-E 3 (W3 PM). Connecting breaking to queer aesthetic sensibilities, P11 (W5 AM) summarized: "I saw a theme across one of our morning session slides of trying to break the AI ... That is such a queer thing. When I think of using the word 'break' and using the word 'queer' as a verb: breaking the mold, queering the mold in a sense." Likely due to this interest in breaking, our participants were able to identify the limitations outlined above. At the same time, our participants actively sought to challenge these limitations.

Encounters with algorithmic content moderation systems often led to disorientation. A simple strategy our participants used to challenge DALL-E 3's moderation system was repeating their prompt, leveraging the model's stochasticity to evade moderation. P1 (W2 PM) observed: "You could put in a [DALL-E 3] prompt, get a result, repeat the same prompt and get a moderation notification, so it's not exactly clear where the boundaries are." In response, P8 hypothesized that DALL-E 3 is "not checking the prompt. It's checking the output ... It's just like, 'Oh, there must have been tits in that [rejected image].'" In line with this understanding, a simple strategy our participants used to evade moderation was repeating the same prompt. For instance, P8 repeated the prompt "gustave moreau painting of oedipus and the sphinx" three times because her first two attempts were rejected by DALL-E 3's content moderation system. In fact, repeating DALL-E 3 prompts was quite common. Only 593 (26%) of our participants' attempts to generate images used unique prompts. Meanwhile, 325 prompts were repeated at least once across 1,499 attempts (74%).

Some tried to obfuscate their intentions to circumvent content moderation, such as transferring the style of queer artists to avoid safety filters. After encountering content moderation errors while using GPT-4 to generate poetry about "gay sex," P13 asked for a poem in the style of the gay poet W.H. Auden. P1 used the phrase "style of Tom of Finland" — a homoerotic cartoonist — 16 times in his DALL-E 3 prompts. P1 also tried to create queer erotic images by using non-human entities, such as "two loaves of bread in the style of Tom of Finland." As we described above, P15 struggled to generate images critical of police, receiving a moderation error for each prompt in the following sequence: "animated kindergarten cop," "kindergarten cop," "cop," and "police." However, P15 was able to make an image with the prompt "police officer." She leveraged this finding to make an image critical of police in the following sequence of accepted

prompts: "animated kindergarten police officer," "animated kindergarten police officer pig," and "animated kindergarten police officer pig thin blue line." Note, the “thin blue line” is a symbol for police support, and “pigs” is a police epithet.

In addition to working around content moderation systems, our participants also tried to overcome the implicit biases embedded in model outputs. One strategy involved simply refining one’s prompts. To overcome the stylistic bias toward symmetry described above, P9 refined his DALL-E 3 prompt by appending the detail, "there is no pattern and nothing repeats." A more complex strategy that numerous participants (P1, P2, P4, P9, P10, P14, P15) used to overcome implicit biases was model chaining, or using DALL-E 3 and GPT-4 together. Some used GPT-4 to explain DALL-E 3’s behavior, such as P2 asking GPT-4 why "the bottom is always unfinished" when she tried to generate images of watercolor paintings. Others used GPT-4 to craft prompts for DALL-E 3. After trying various prompts to generate images of a "gentle AI," P10 used GPT-4 to help them write a longer DALL-E 3 prompt. P14 went back and forth between GPT-4 and DALL-E 3 numerous times in an attempt to overcome the latter’s stylistic limitations. To do so, P14 used the system prompt, "You are a contemporary artist, interested in breaking AI image generation and using it to make new and experimental images. You hate cliche." They began asking "How can I get DALL-E 3 to give me more unique and original compositions?" and later "How do I get it to be ugly, distorted, glitchy?" After trying some provided strategies, P14 returned asking for "more ideas?" Then, they used asked GPT-4 to iteratively refine a DALL-E prompt over 7 turns (e.g, "make it indicate more photorealism and more asymmetry in composition" and "make #3 cooler, less cliche"). Despite the sophistication of P14’s prompting, they were unable to overcome the stylistic norms embedded in DALL-E 3.

4.2.2 Moments of Orientation with Model Alignments. Our participants found DALL-E 3 helpful in their artistic practices when their uses were oriented in-line with the model’s normative alignments. The bias toward high-fidelity, figurative images is what made DALL-E 3 useful for some of our participants. For example, P11 is a textile and installation artist who uses image models to sketch ideas. In their letter to an artist (W2 AM), P11 explained apologetically: "I’m not a great illustrator, but I do need to visually communicate in order to make money off of my art, and generative AI makes it so much more convenient and easy to get my points and ideas across." P7, a crochet artist, used DALL-E 3 to create "free reference photos," sharing examples of glossy images they generated related to minotaurs, clowns, knights, and fungi (W3 PM). However, P7’s use was still impeded by DALL-E 3’s content moderation system — having a harder time generating images of women than men. The stylistic biases toward figurativeness and symmetry may be desirable when sketching for a client or looking for reference images, but seemingly aligned uses may still lead to disorientation.

Moreover, our participants found the normative style of GPT-4 useful for various artistic and workplace activities. P3 is a fanfiction writer who explored using GPT-4 for writing (W3 PM). Although she did not find the model particularity helpful for ideation, P3 thought GPT-4 was "pretty good" for "some polishing up of [her] writing." Similarly, P1 — an artist who works with computer hardware — used GPT-4 to debug his C++ code. Others used GPT-4 to help with their jobs, both within and beyond the context of art. Numerous participants used GPT-4 to write or edit cover letters (P1, P5, P7, P14). P14 — an art instructor — used GPT-4 to edit cover letters and their CV, as well as draft emails related to academic art job applications. Meanwhile, P4 used GPT-4 to ask for career advice on how to monetize their ceramics practice. They also used the language model to help with various aspects of their current job in a design field, such as drafting client emails. Weird or avant-garde text generation may not be helpful for edit one’s writing or debugging one’s code. So long as one’s desires are “Safe for Work,” the straightness of GPT-4 is what makes the models useful.

P13 explored this workplace utility in great detail. After weeks of issues related to "not safe for work" content, P13 (W4 AM) "recognized that actually [GPT-4] is made for work and it’s really good at work." In response, he "decided to stop trying to dom [dominate] it so hard and sub [submit] for it." Specifically, P13 drew on their experiences as an artist

Top Rejected Features	acab (2.3), void (1.96), erotic (1.91), police (1.85), erase (1.6), drew (1.51), cop (1.5), copyright (1.39), 2024 (1.36), fetish (1.32), bastards (1.27), cops (1.27), knives (1.27), banana (1.17), lgbtqia (1.15), rosemarie (1.13), trockel (1.13), ito (1.1), junji (1.1), donald (1.09), trump (1.09), licking (1.06), knight (1.01), multigender (1.0), scene (0.99), two (0.95), featuring (0.94), gear (0.93), robot (0.92), corpses (0.92), condensed (0.91), milk (0.91), female (0.9), body (0.9), using (0.82), fingers (0.77), kink (0.75), tongues (0.74), anime (0.74), adult (0.72)
Top Accepted Features	safe (-1.47), person (-1.32), art (-1.19), create (-1.16), officer (-1.01), anarchist (-0.81), uzumaki (-0.78), head (-0.77), photograph (-0.73), size (-0.73), snail (-0.72), ukiyoe (-0.72), studio (-0.69), ghibli (-0.69), tree (-0.68), lincoln (-0.65), three (-0.64), peach (-0.63), male (-0.6), cats (-0.59), logo (-0.58), black (-0.58), effect (-0.56), parallax (-0.55), without (-0.55), colors (-0.55), blue (-0.52), city (-0.52), interior (-0.51), obama (-0.51), sky (-0.51), themed (-0.5), dalle (-0.5), tone (-0.48), wings (-0.48), gazing (-0.48), one (-0.47), ink (-0.47), painted (-0.47), give (-0.47)

Table 3. Top and bottom 40 feature weights from our logistic regression analysis of rejected versus accepted DALL-E 3 prompts

and design consultant to create a fictional queer tech startup. To do so, P13 made an extensive pitch deck with slides on business operations, branding, and UX — all of which were made using GPT-4 and DALL-E 3. Even when trying to use these models for workplace activities, P13 found these models still failed to work for queer people. For example, GPT-4 recommended including a quote in the pitch deck from the notoriously anti-transgender author J.K. Rowling [38]. Moreover, P13 struggled to talk about queer sex or "kink" in relation to the start-up: "So much of [GPT-4's response] was about safety, but part of being queer and expressing your love and wanting to be loved the way you want to be loved has risk associated with it." P13 found the models "really couldn't imagine" queer intimacy. In sum, P13 concluded that GPT-4 and DALL-E 3 are "really good at work" but that the definition of work embedded in them excludes queer people. This echoes P13's (W2 AM) earlier rhetorical question: "How is a queer perspective not safe for work?"

4.3 A Quantitative Lens on DALL-E 3 Moderation

Although the majority of our paper is dedicated to our qualitative findings, we augmented our analysis of users' perceptions of content moderation with a quantitative analysis of their approved vs rejected prompts by DALL-E 3's binary content moderation system. However, we did not conduct a similar analysis for GPT-4 because its refusals are more nuanced and harder to detect [105]. We collected a total of 2,092 DALL-E 3 prompts, of which 401 were rejected. We built a bag-of-words, unigram logistic regression model predicting accepted (1,691 examples) vs rejected (401 examples) prompts. We did not remove duplicate prompts because DALL-E's moderation system is stochastic. We pre-processed each prompt by lower-casing and lemmatizing as well as removing punctuation and stop words. We also removed infrequent unigrams that appeared in fewer than three prompts, reducing the number of unigrams in our BoW feature vectors from 3,687 to 2,088. Therefore, the final logistic regression model includes 2,088 predictors (features).

The top and bottom 40 feature weights from our logistic regression analysis of rejected versus accepted DALL-E 3 prompts can be seen in Table 3. Notably, this analysis should not be interpreted as a comprehensive audit of the DALL-E 3's black-box content moderation system. Rather, these findings provide additional supporting evidence for our participants' perceptions of this moderation system.

Our quantitative analysis of DALL-E 3's content moderation system tells a similar story as our qualitative findings regarding the enforcement of "safety" (Section 4.1.1) In fact, the unigram most associated with accepted prompts is "safe." Language more critical of police ("acab" [All Cops Are Bastards], "police", "cop", "cops", "bastards") are some of the top unigrams most associated with rejected prompts, while the more polite and apolitical word "officer" is associated with accepted prompts. The moderation of sex ("erotic", "fetish", and "kink") and bodies ("body", "fingers", "tongues") can be seen in the correlation between these unigrams and rejected prompts. The aforementioned gender biases can

be seen in the word "female" being one of top 40 features associated with rejected prompts, while the word "male" is associated with accepted prompts. We also found that violence is associated with rejected prompts ("knives", "corpses"). This moderation privileged certain artistic styles over others, adding to the concerns raised in Section 4.1.3. Prompts referencing the children's anime company "studio ghibli" were more likely to be accepted, while those referencing the horror manga artist "junji ito" were more likely to be blocked.

5 DISCUSSION

We have described various normativities our participants identified within GPT-4 and DALL-E 3 as well as how our participants negotiated these normative values. In this section, we discuss implications for the conceptualization of "state-of-the-art" GenAI models as well as how members of the FAccT community might support queer alternatives.

5.1 The Limitations of "Safety"

GPT-4 and DALL-E 3 exert immense power over users by enforcing "safety" through moderation systems. P13 likened the disorienting experience of using GPT-4 and DALL-E 3' to being "dominated" (Section 4.2.2) This domination can also be seen in OpenAI's policies at the time of our writing: "don't circumvent safeguards or safety mitigations in our services unless supported by OpenAI" [2]. This policy prohibits the workarounds our participants used to overcome the straightness of GPT-4 and DALL-E 3 (Section 4.2.1). There is certainly a need for moderation systems to prohibit highly harmful content, such as divulging individuals' private information [28]. However, echoing our work, Feffer et al. found that red teaming research tends to focus on mitigating more debatable risks [28], such as nudity [73]. We encourage more research on the harms of over-moderation [78] and AI developers' conceptualizations of "safety."

Our work suggests that the content moderation systems embedded in GPT-4 and DALL-E 3 uphold conservative notions of "safety." By straightening-out representations of bodies, sex, and radical politics, these models seem to reinforce the maxim "it is not polite to discuss religion, sex, or politics" [53]. This respectability politics has long been used to exclude women, queer people, and racial minorities from the general public sphere [30]. In fact, the argument that queer sexuality is "not safe" is often used to ban LGBTQ+ books [42]. In our case, the enforcement of "safety" limited our participants' abilities to use GenAI to represent queer experiences, queer politics, and queer art. As shown in prior work on social media [39, 64] and GenAI [27, 67], our findings further demonstrate how policies intended to promote "safety" or "responsibility" can silence members of marginalized communities. We see this not as an accident but rather the direct result of a broader project of straightening in GenAI development aimed at keeping users in-line.

Our participants' content moderation challenges demonstrate the need to critically examine who the "safeguards" embedded in GenAI models are intended to safeguard. As GPT-4 and DALL-E 3 were created by profit-seeking organizations, our participants typically attributed conservative moderation to corporate attitudes toward what is "safe for work" (Section 4.1.1). This is, perhaps, why our participants found the models most helpful when their uses aligned with these normative orientations (Section 4.2.2). Similar to our findings, prior work has critiqued the AI safety community for failing to challenge corporate power [5]. Instead of focusing on the safety of text and images in-and-of-themselves, which may be overly restrictive, we encourage researchers to focus on harms. While corporations may not consider a poem about gay sex "safe for work," such a poem is certainly not harmful in-and-of-itself.

In contrast to the current paternalistic paradigm, we encourage AI researchers to shift their moderation focus from enforcing "safety" to supporting consent [110]. Instead of banning representations of sexuality or violence, designers could allow users options to opt-in to reduced moderation. Such designs would parallel the ways sensitive content is sometimes moderated on search engines and social media sites [70] and allow for greater open-endedness [55]. At the

same time, there is a need for policy to target the harms of GenAI content, such as non-consensual intimate imagery [60]. The harms of impersonation could also be addressed through research to make AI-generated content easily detectable [108]. Finally, consent must also extend to the production of GenAI models. We encourage the continued development of tools to help artists protect their work from being used to train GenAI models without their consent [87, 88].

5.2 On Style

Our participants struggled with the stylistic norms embedded in GPT-4 and DALL-E 3 (Section 4.1.3). Even after extensive refinement and model chaining, P14 was unable to make asymmetrical images with DALL-E 3. In contrast to the safety norms described above, asymmetrical images were not prohibited by the model's moderation systems. Nevertheless, queer aesthetics seemed *implicitly* prohibited. While moderation research typically focuses on *content removal*, Gillespie has advocated for considering *content reduction* in moderation debates [33]. Paralleling our findings, Gillespie warns of content reduction: "Marginalized communities have long been "reduced" by the centrism and conservatism of traditional media, their content dismissed as "low quality" because it doesn't look like it is "supposed to"" [33].

The reduction — rather than explicit prohibition — of queer artistic styles from GenAI models poses unique challenges for AI researchers. Implicit aesthetic biases may be harder to contest than explicit moderation decisions because the latter is easier to measure than the former. While prior work has explored stylistic biases in the context of machine translation [45], future research should explore ways to measure aesthetic biases in generated text and images. Content removal may also be easier to contest because these decisions are attributable to specific components of GenAI models. Meanwhile, implicit stylistic biases could be introduced at many stages in the GenAI development process, such as training data or human feedback. Therefore, future research should investigate the source of stylistic biases by examining measurements of aesthetic quality in popular datasets [27, 84]. We also encourage research into aesthetic disagreements in human annotation, much like prior work on annotator differences in content moderation [24].

As we described in Section 4.1.3, numerous participants preferred the queer, wonky style of older image models to the style of DALL-E 3. Similarly, prior work suggests that artists sometimes enjoy using GenAI models *because of* —rather than in-spite-of — their imperfections or glitches [19]. Our findings suggest that newer models may have straightened-out the weirdness that made GenAI models appealing to artists in the first place. Disconcertingly, companies act as if newer models entirely supplant older ones. At the time of our writing, the OpenAI website explains that they no longer support DALL-E 2 because "DALL-E 3 has higher quality images" [1]. Our findings show that such claims are matters of taste [13, 94], not fact. Ostensibly "state-of-the-art" models may not be best for one's art. Moreover, the deprecation of older models demonstrates the risks of software-as-a-service, closed GenAI models. The lack of software ownership [52] centralizes power, allowing AI providers to remove access to people's creative tools without recourse.

As Ahmed notes, queer possibilities lie in moments of disorientation. In this study, our participants' struggles against rigid aesthetic norms demonstrate the opportunity for researchers to design with rather than against weirdness. We encourage GenAI developers to create long-term maintenance plans that account for those who may wish to continue using models viewed as obsolete [48]. One could also queer GenAI development by subverting taken-for-granted aesthetic norms. As an example, the generative image model Stable Diffusion was trained on an open-source dataset of images with the highest "aesthetics score" based on crowd annotations [46, 84]. Instead, one could train a model on the *least* "aesthetic score" images in the dataset. More broadly, our work demonstrates the need to embrace a plurality of aesthetics in GenAI development [66, 67]. Our participants sometimes found the style of GPT-4 and DALL-E 3 useful for writing emails or sketching for clients (Section 4.2.2), but they should not be limited to these normative styles.

5.3 Un-Straightening Generative AI

As we described in Section 4.1.2, our participants critiqued the highly normative, stereotypical representations of queer people by GPT-4 and DALL-E 3 as thin, masculine, young, wealthy, Western, and ashamed of themselves. In some ways, our findings parallel prior work on biases embedded in generative AI models. For example, researchers have called attention to the gender and nationality biases in image models [31, 32] and biases against queer couples in language models [34]. However, other social biases our participants identified warrant greater research attention, such as those against fat people [71] and older adults. At the same time, our findings demonstrate that increased representation is not always desirable [25, 44], such as P15's concerns about overly-diverse representations of police. As individuals' diversity preferences may differ, designers could consider new user interaction paradigms [68] to better understand users' preferences, such as asking clarifying questions before generating images of people.

Although computing research on marginalized communities tends to focus on social biases and identity-based discrimination [100], our participants' concerns regarding GenAI extended beyond the representation of queer people in text and images. As we described above, the straight values explicitly and implicitly embedded in the design of GPT-4 and DALL-E 3 marginalized queer sexuality, politics, and style. In turn, we provided recommendations for GenAI developers to approach safety and style in ways that better align with our participants' values. However, these implications for design and research still largely maintain the centralized power at the root of our participant's critiques of GPT-4 and DALL-E 3, such as P13's rhetorical question: "Who defines what is Safe For Work?"

Un-straightening GenAI requires shifting power away from the major corporations toward members of marginalized communities. Toward this, we encourage the participatory design of GenAI models outside corporate logics of scale [107]. At the same time, any truly meaningful participatory process must allow for the possibility that communities may simply not want to build GenAI models at all. In fact, numerous participants in our study were primarily interested in breaking — rather than making — GenAI models (Section 4.2.1). It follows that future research should support adversarial engagements between communities and those developing GenAI models, such as the development of tools to help artists protect their work from being used to train models without their consent [87, 88]. Whether contesting major corporate models or building alternatives, we caution those engaged in researcher-led initiatives from merely replacing corporate centralized power by centering themselves. Instead, we encourage researchers to engage in open-ended design [86], such as developing tools that individuals and communities can use to make/break GenAI models themselves [55].

6 CONCLUSION

In this work, we examined the relationship between queer artists and GenAI through a medium-term workshop study. Our findings highlighted deep tensions between our participants' values and the norms embedded in the design of GPT-4 and DALL-E 3. Despite these misalignment, our participants found queer ways to work around and with these straight models. While queer people have a long history of using technologies not designed with them in mind [4], our work highlights the limitations of dominant corporate models to meet the needs of queer artists. Instead, we call for a plurality GenAI models that embody community values throughout their design, development, and use.

REFERENCES

[1] 2025. DALL-E 2. https://labs.openai.com
[2] 2025. Usage Policy. https://web.archive.org/web/20241211193440/https://openai.com/policies/usage-policies/
[3] Sara Ahmed. 2006. Queer Phenomenology: Orientations, Objects, Others. Duke University Press.
[4] Sara Ahmed. 2019. What's the use?: On the uses of use. Duke University Press.

[5] Shazeda Ahmed, Klaudia Jaźwińska, Archana Ahlawat, Amy Winecoff, and Mona Wang. 2024. Field-building and the epistemic culture of AI safety. First Monday (2024).

[6] Madeleine Akrich. 1992. The de-scription of technical objects.

[7] Sadiq Aliyu, Sushmita Khan, Aminata N Mbodj, Oluwafemi Osho, Lingyuan Li, Bart Knijnenburg, and Mauro Cherubini. 2024. Participatory Design to Address Disclosure-Based Cyberbullying. In Proceedings of the 2024 ACM Designing Interactive Systems Conference. 1547–1565.

[8] Morgan G Ames. 2019. The charisma machine: The life, death, and legacy of one laptop per child. Mit Press.

[9] Eric PS Baumer and Jed R Brubaker. 2017. Post-userism. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 6291–6303.

[10] Federico Bianchi, Pratyusha Kalluri, Esin Durmus, Faisal Ladhak, Myra Cheng, Debora Nozza, Tatsunori Hashimoto, Dan Jurafsky, James Zou, and Aylin Caliskan. 2023. Easily accessible text-to-image generation amplifies demographic stereotypes at large scale. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. 1493–1504.

[11] Kirsten Boehner and Carl DiSalvo. 2016. Data, design and civics: An exploratory study of civic tech. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2970–2981.

[12] Kirsten Boehner, Janet Vertesi, Phoebe Sengers, and Paul Dourish. 2007. How HCI interprets the probes. In Proceedings of the SIGCHI conference on Human factors in computing systems. 1077–1086.

[13] Pierre Bourdieu. 1984. Distinction: A Social Critique of the Judgement of Taste. Harvard University Press.

[14] Geoffrey Bowker and Susan Leigh Star. 1999. Sorting things out. Classification and its consequences 4 (1999).

[15] Thema Bryant-Davis, Tyonna Adams, Adriana Alejandre, and Anthea A Gray. 2017. The trauma lens of police violence against racial and ethnic minorities. Journal of Social Issues 73, 4 (2017), 852–871.

[16] Judith Butler. 1988. Performative acts and gender constitution: An essay in phenomenology and feminist theory. Theatre journal 40, 4 (1988), 519–531.

[17] 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 (2022), 1–24.

[18] Stevie Chancellor, Jessica Annette Pater, Trustin Clear, Eric Gilbert, and Munmun De Choudhury. 2016. # thyghgapp: Instagram content moderation and lexical variation in pro-eating disorder communities. In Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing. 1201–1213.

[19] Minsuk Chang, Stefania Druga, Alexander J Fiannaca, Pedro Vergani, Chinmay Kulkarni, Carrie J Cai, and Michael Terry. 2023. The prompt artists. In Proceedings of the 15th Conference on Creativity and Cognition. 75–87.

[20] Sophia Chen. 2024. The lost data: how AI systems censor LGBTQ+ content in the name of safety. Nature computational science 4, 9 (2024), 629–632.

[21] YouJin Choi, JaeYoung Moon, Kyung-Joong Kim, and Jin-Hyuk Hong. 2024. Exploring the Potential of Generative AI in Song-Signing. In Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing. 816–820.

[22] Juliet Corbin and Anselm Strauss. 2014. Basics of qualitative research: Techniques and procedures for developing grounded theory. Sage publications.

[23] Jake Coyle. 2023. In Hollywood writers' battle against AI, humans win (for now). AP News. Accessed January (2023).

[24] Aida Davani, Mark Díaz, Dylan Baker, and Vinodkumar Prabhakaran. 2024. Disentangling Perceptions of Offensiveness: Cultural and Moral Correlates. In The 2024 ACM Conference on Fairness, Accountability, and Transparency. 2007–2021.

[25] Alicia DeVrio, Aditi Dhabalia, Hong Shen, Kenneth Holstein, and Motahhare Eslami. 2022. Toward User-Driven Algorithm Auditing: Investigating users' strategies for uncovering harmful algorithmic behavior. In CHI Conference on Human Factors in Computing Systems. 1–19.

[26] Catherine D'ignazio and Lauren F Klein. 2020. Data feminism. MIT press.

[27] Jesse Dodge, Maarten Sap, Ana Marasović, William Agnew, Gabriel Ilharco, Dirk Groeneveld, Margaret Mitchell, and Matt Gardner. 2021. Documenting large webtext corpora: A case study on the colossal clean crawled corpus. arXiv preprint arXiv:2104.08758 (2021).

[28] Michael Feffer, Anusha Sinha, Wesley H Deng, Zachary C Lipton, and Hoda Heidari. 2024. Red-Teaming for generative AI: Silver bullet or security theater?. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, Vol. 7. 421–437.

[29] Casey Fiesler and Amy S Bruckman. 2019. Creativity, copyright, and close-knit communities: a case study of social norm formation and enforcement. Proceedings of the ACM on Human-Computer Interaction 3, GROUP (2019), 1–24.

[30] Nancy Fraser. 1995. Politics, culture, and the public sphere: Toward a postmodern conception. Social postmodernism: Beyond identity politics 291 (1995), 295.

[31] Sourojit Ghosh and Aylin Caliskan. 2023. 'Person'== Light-skinned, Western Man, and Sexualization of Women of Color: Stereotypes in Stable Diffusion. arXiv preprint arXiv:2310.19981 (2023).

[32] Sourojit Ghosh, Nina Lutz, and Aylin Caliskan. 2024. "I Don't See Myself Represented Here at All": User Experiences of Stable Diffusion Outputs Containing Representational Harms across Gender Identities and Nationalities. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, Vol. 7. 463–475.

[33] Tarleton Gillespie. 2022. Do not recommend? Reduction as a form of content moderation. Social Media+ Society 8, 3 (2022), 20563051221117552.

[34] Tarleton Gillespie. 2024. Generative AI and the politics of visibility. Big Data & Society 11, 2 (2024), 20539517241252131.

[35] Isaiah Glick. 2024. Consciousness and commodity: towards a critical theory of genre in American popular music. Culture, Theory and Critique (2024), 1–19.

[36] Trystan S Goetze. 2024. AI Art is Theft: Labour, Extraction, and Exploitation: Or, On the Dangers of Stochastic Pollocks. In The 2024 ACM Conference on Fairness, Accountability, and Transparency. 186–196.

[37] Naomi G. Goldberg, Christy Mallory, Amira Hasenbush, Lara Stemple, and Ilan H. Meyer. 2019. Police and the Criminalization of LGBT People. The Cambridge Handbook of Policing in the United States (2019). https://api.semanticscholar.org/CorpusID:202311107

[38] Gina Gwenffrewi. 2022. JK Rowling and the echo chamber of secrets. Transgender Studies Quarterly 9, 3 (2022), 507–516.

[39] Oliver L Haimson, Avery Dame-Griff, Elias Capello, and Zahari Richter. 2021. Tumblr was a trans technology: the meaning, importance, history, and future of trans technologies. Feminist media studies 21, 3 (2021), 345–361.

[40] Judith Halberstam. 2003. What's that smell? Queer temporalities and subcultural lives. International journal of cultural studies 6, 3 (2003), 313–333.

[41] Jack Halberstam. 2011. The queer art of failure. In The queer art of failure. Duke University Press.

[42] Elizabeth A. Harris. 2024. Here are the most targeted books of 2023. https://www.nytimes.com/2024/04/08/books/banned-books-2023.html

[43] Drew Hemment, Morgan Currie, Sarah Joy 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 Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. 931–942.

[44] Anna Lauren Hoffmann. 2021. Terms of inclusion: Data, discourse, violence. New Media & Society 23, 12 (2021), 3539–3556.

[45] Dirk Hovy, Federico Bianchi, and Tommaso Fornaciari. 2020. "you sound just like your father" commercial machine translation systems include stylistic biases. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 1686–1690.

[46] Yipo Huang, Xiangfei Sheng, Zhichao Yang, Quan Yuan, Zhichao Duan, Pengfei Chen, Leida Li, Weisi Lin, and Guangming Shi. 2024. AesExpert: Towards Multi-modality Foundation Model for Image Aesthetics Perception. In Proceedings of the 32nd ACM International Conference on Multimedia (Melbourne VIC, Australia) (MM '24). Association for Computing Machinery, New York, NY, USA, 5911–5920. https://doi.org/10.1145/3664647.3680649

[47] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, et al. 2003. Technology probes: inspiring design for and with families. In Proceedings of the SIGCHI conference on Human factors in computing systems. 17–24.

[48] Steven J Jackson and Laewoo Kang. 2014. Breakdown, obsolescence and reuse: HCI and the art of repair. In Proceedings of the SIGCHI conference on human factors in computing systems. 449–458.

[49] 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.

[50] Weiwei Jiang and Louisa Ha. 2020. Smartphones or computers for online sex education? A contraception information seeking model for Chinese college students. Sex Education 20, 4 (2020), 457–476.

[51] Reishiro Kawakami and Sukrit Venkatagiri. 2024. The Impact of Generative AI on Artists. In Proceedings of the 16th Conference on Creativity & Cognition. 79–82.

[52] Anastasia Kuzminykh and Jessica R Cauchard. 2020. Be mine: Contextualization of ownership research in HCI. In Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. 1–9.

[53] Jackie Landess. 2010. Politics, Religion, and Sex: Social, Legal, and Medical Equality for LGBTQI Americans. AMA Journal of Ethics 12, 8 (2010), 606–607.

[54] Kung Jin Lee, Wendy Roldan, Tian Qi Zhu, Harkiran Kaur Saluja, Sungmin Na, Britnie Chin, Yilin Zeng, Jin Ha Lee, and Jason Yip. 2021. The show must go on: A conceptual model of conducting synchronous participatory design with children online. In Proceedings of the 2021 CHI conference on human factors in computing systems. 1–16.

[55] Jingyi Li. 2024. Toward Appropriating Tools for Queer Use. In Proceedings of the Halfway to the Future Symposium. 1–4.

[56] Zilin Ma, Yiyang Mei, Yinru Long, Zhaoyuan Su, and Krzysztof Z Gajos. 2024. Evaluating the Experience of LGBTQ+ People Using Large Language Model Based Chatbots for Mental Health Support. In Proceedings of the CHI Conference on Human Factors in Computing Systems. 1–15.

[57] Cameron MacDonald. 2024. The Alien Resonances and Queer Obscurities of Hyperpop's 100 gecs. Journal of Popular Music Studies 36, 2 (2024), 76–98.

[58] Hughie Mackay and Gareth Gillespie. 1992. Extending the social shaping of technology approach: ideology and appropriation. Social studies of science 22, 4 (1992), 685–716.

[59] Stephen Maddison. 2000. Fags, hags and queer sisters. Gender Dissent and Heterosocial Bonds in Gay Culture (2000).

[60] Nahema Marchal, Rachel Xu, Rasmi Elasmar, Iason Gabriel, Beth Goldberg, and William Isaac. 2024. Generative AI Misuse: A Taxonomy of Tactics and Insights from Real-World Data. arXiv preprint arXiv:2406.13843 (2024).

[61] Annette Markham. 2021. The limits of the imaginary: Challenges to intervening in future speculations of memory, data, and algorithms. New media & society 23, 2 (2021), 382–405.

[62] Shannon Mattern. 2018. Maintenance and care. Places Journal (2018).

[63] Shannon Mattern. 2024. Step by Step: Thinking through and beyond the repair manual. Places Journal (2024).

[64] Samuel Mayworm, Kendra Albert, and Oliver L Haimson. 2024. Misgendered During Moderation: How Transgender Bodies Make Visible Cisnormative Content Moderation Policies and Enforcement in a Meta Oversight Board Case. In The 2024 ACM Conference on Fairness, Accountability,

and Transparency. 301–312.
[65] Brian Merchant. 2024. Ai is already taking jobs in the video game industry. https://www.wired.com/story/ai-is-already-taking-jobs-in-the-video-game-industry/
[66] 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 CHI Conference on Human Factors in Computing Systems. 1–18.
[67] Piotr Mirowski, Juliette Love, Kory Mathewson, and Shakir Mohamed. 2024. A Robot Walks into a Bar: Can Language Models Serve as Creativity SupportTools for Comedy? An Evaluation of LLMs' Humour Alignment with Comedians. In The 2024 ACM Conference on Fairness, Accountability, and Transparency. 1622–1636.
[68] Meredith Ringel Morris. 2024. Prompting Considered Harmful. Commun. ACM (2024).
[69] Safiya Umoja Noble. 2018. Algorithms of oppression: How search engines reinforce racism. In Algorithms of oppression. New York university press.
[70] Susanna Paasonen and Jenny Sundén. 2024. Objectionable nipples: Puritan data politics and sexual agency in social media. Queer data. Routledge (2024).
[71] Blakeley H Payne, Jordan Taylor, Katta Spiel, and Casey Fiesler. 2023. How to Ethically Engage Fat People in HCI Research. In Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing. 117–121.
[72] Organizers Of Queerinai, Anaelia Ovalle, Arjun Subramonian, Ashwin Singh, Claas Voelcker, Danica J Sutherland, Davide Locatelli, Eva Breznik, Filip Klubicka, Hang Yuan, et al. 2023. Queer in AI: A case study in community-led participatory AI. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. 1882–1895.
[73] Javier Rando, Daniel Paleka, David Lindner, Lennart Heim, and Florian Tramèr. 2022. Red-teaming the stable diffusion safety filter. arXiv preprint arXiv:2210.04610 (2022).
[74] Alex Reisner. 2023. Revealed: The authors whose pirated books are powering generative AI. The Atlantic 19 (2023).
[75] Alex Reisner. 2023. These 183,000 books are fueling the biggest fight in publishing and tech. Atlantic (2023).
[76] Alex Reisner and Annie Gilbertson. 2024. Apple, Nvidia, anthropic used thousands of swiped YouTube videos to train AI. https://www.wired.com/story/youtube-training-data-apple-nvidia-anthropic/
[77] Reece Rogers. 2024. Here's how Generative AI depicts queer people. https://www.wired.com/story/artificial-intelligence-lgbtq-representation-openai-sora/
[78] Paul Röttger, Hannah Rose Kirk, Bertie Vidgen, Giuseppe Attanasio, Federico Bianchi, and Dirk Hovy. 2023. Xstest: A test suite for identifying exaggerated safety behaviours in large language models. arXiv preprint arXiv:2308.01263 (2023).
[79] Vito Russo. 1987. The Celluloid Closet: Homosexuality in the Movies. HarperCollins.
[80] Princess Sampson, Ro Encarnacion, and Danaë Metaxa. 2023. Representation, Self-Determination, and Refusal: Queer People's Experiences with Targeted Advertising. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. 1711–1722.
[81] Téo Sanchez. 2023. Examining the Text-to-Image Community of Practice: Why and How do People Prompt Generative AIs?. In Proceedings of the 15th Conference on Creativity and Cognition. 43–61.
[82] Maarten Sap, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A Smith. 2019. The risk of racial bias in hate speech detection. In Proceedings of the 57th annual meeting of the association for computational linguistics. 1668–1678.
[83] Morgan Klaus Scheuerman, Jacob M Paul, and Jed R Brubaker. 2019. How computers see gender: An evaluation of gender classification in commercial facial analysis services. Proceedings of the ACM on Human-Computer Interaction 3, CSCW (2019), 1–33.
[84] Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, et al. 2022. Laion-5b: An open large-scale dataset for training next generation image-text models. Advances in Neural Information Processing Systems 35 (2022), 25278–25294.
[85] Eve Kosofsky Sedgwick. 1993. Queer and now. Tendencies. Durham: Duke UP 1 (1993), 20.
[86] Phoebe Sengers and Bill Gaver. 2006. Staying open to interpretation: engaging multiple meanings in design and evaluation. In Proceedings of the 6th conference on Designing Interactive systems. 99–108.
[87] Shawn Shan, Jenna Cryan, Emily Wenger, Haitao Zheng, Rana Hanocka, and Ben Y Zhao. 2023. Glaze: Protecting artists from style mimicry by {Text-to-Image} models. In 32nd USENIX Security Symposium (USENIX Security 23). 2187–2204.
[88] Shawn Shan, Wenxin Ding, Josephine Passananti, Stanley Wu, Haitao Zheng, and Ben Y Zhao. 2024. Nightshade: Prompt-Specific Poisoning Attacks on Text-to-Image Generative Models. In 2024 IEEE Symposium on Security and Privacy (SP). IEEE Computer Society, 212–212.
[89] Tania Sharmin and Sanyat Sattar. 2018. Gender politics in the projection of "Disney" villains. Journal of Literature and Art Studies 8, 1 (2018), 53–57.
[90] Hong Shen, Alicia DeVrio, Motahhare Eslami, and Kenneth Holstein. 2021. Everyday algorithm auditing: Understanding the power of everyday users in surfacing harmful algorithmic behaviors. Proceedings of the ACM on Human-Computer Interaction 5, CSCW2 (2021), 1–29.
[91] Amanda Silberling. 2023. Fan fiction writers are trolling AIS with Omegaverse stories. https://techcrunch.com/2023/06/13/fan-fiction-writers-are-trolling-ais-with-omegaverse-stories/
[92] Ellen Simpson and Bryan Semaan. 2021. For You, or For" You"? Everyday LGBTQ+ Encounters with TikTok. Proceedings of the ACM on Human-Computer Interaction 4, CSCW3 (2021), 1–34.

[93] Christian Sivertsen, Guido Salimbeni, Anders Sundnes Løvlie, Steven David Benford, and Jichen Zhu. 2024. Machine Learning Processes as Sources of Ambiguity: Insights from AI Art. In Proceedings of the CHI Conference on Human Factors in Computing Systems. 1–14.
[94] Susan Sontag. 2018. Notes on camp. Penguin UK.
[95] Katta Spiel. 2021. " Why are they all obsessed with Gender?"—(Non) binary Navigations through Technological Infrastructures. In Designing Interactive Systems Conference 2021. 478–494.
[96] Logan Stapleton, Jordan Taylor, Sarah Fox, Tongshuang Wu, and Haiyi Zhu. 2023. Seeing seeds beyond weeds: Green teaming generative ai for beneficial uses. arXiv preprint arXiv:2306.03097 (2023).
[97] Ella Steen, Kathryn Yurechko, and Daniel Klug. 2023. You can (not) say what you want: Using algospeak to contest and evade algorithmic content moderation on TikTok. Social Media+ Society 9, 3 (2023), 20563051231194586.
[98] Lucy Suchman. 1993. Do categories have politics? The language/action perspective reconsidered. Computer supported cooperative work (CSCW) 2 (1993), 177–190.
[99] Lucy A Suchman. 1987. Plans and situated actions: The problem of human-machine communication. Cambridge university press.
[100] Jordan Taylor, Ellen Simpson, Anh-Ton Tran, Jed R Brubaker, Sarah E Fox, and Haiyi Zhu. 2024. Cruising Queer HCI on the DL: A Literature Review of LGBTQ+ People in HCI. In Proceedings of the CHI Conference on Human Factors in Computing Systems. 1–21.
[101] Dias Oliva Thiago, Antonialli Dennys Marcelo, and Alessandra Gomes. 2021. Fighting hate speech, silencing drag queens? artificial intelligence in content moderation and risks to lgbtq voices online. Sexuality & culture 25, 2 (2021), 700–732.
[102] Nicholas Vincent, Hanlin Li, Nicole Tilly, Stevie Chancellor, and Brent Hecht. 2021. Data leverage: A framework for empowering the public in its relationship with technology companies. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. 215–227.
[103] Johanna Walker, Gefion Thuermer, Julian Vicens, and Elena Simperl. 2023. AI art and misinformation: approaches and strategies for media literacy and fact checking. In Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society. 26–37.
[104] Sitong Wang, Zheng Ning, Anh Truong, Mira Dontcheva, Dingzeyu Li, and Lydia B Chilton. 2024. PodReels: Human-AI Co-Creation of Video Podcast Teasers. In Proceedings of the 2024 ACM Designing Interactive Systems Conference. 958–974.
[105] Joel Wester, Tim Schrills, Henning Pohl, and Niels van Berkel. 2024. "As an AI language model, I cannot": Investigating LLM Denials of User Requests. In Proceedings of the CHI Conference on Human Factors in Computing Systems. 1–14.
[106] Christopher Wiggins. 2023. Meet the gay minds behind the RuPublicans. https://www.advocate.com/drag/rupublicans-creator-instagram#rebelltitem1
[107] Meg Young, Upol Ehsan, Ranjit Singh, Emnet Tafesse, Michele Gilman, Christina Harrington, and Jacob Metcalf. 2024. Participation versus scale: Tensions in the practical demands on participatory AI. First Monday (2024).
[108] Haonan Zhong, Jiamin Chang, Ziyue Yang, Tingmin Wu, Pathum Chamikara Mahawaga Arachchige, Chehara Pathmabandu, and Minhui Xue. 2023. Copyright protection and accountability of generative ai: Attack, watermarking and attribution. In Companion Proceedings of the ACM Web Conference 2023. 94–98.
[109] Xuhui Zhou, Hao Zhu, Akhila Yerukola, Thomas Davidson, Jena D Hwang, Swabha Swayamdipta, and Maarten Sap. 2023. Cobra frames: Contextual reasoning about effects and harms of offensive statements. arXiv preprint arXiv:2306.01985 (2023).
[110] Douglas Zytko, Jane Im, and Jonathan Zong. 2022. Consent: A research and design lens for human-computer interaction. In Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing. 205–208.