The Heteronormative Male Gaze: Experiences of Sexual Content Moderation among Queer Instagram Users in Berlin

RACHEL GRIFFIN[♦1]
Sciences Po Law School, France

Recent research suggests that social media moderation policies banning sexual content systematically disfavor LGBTQIA+ people. However, specific impacts on particular communities remain understudied. This article contributes to the literature with a qualitative interview-based study of LGBTQIA+ Instagram users in Berlin. Participants perceived Instagram's policies as not only homophobic, but heteronormative in the broadest sense, enforcing interrelated norms around gender, sexuality, and bodies that privilege cis-straight male perspectives. These policies significantly limited queer self-expression. Importantly, these impacts were largely relational: Users were not only affected by the moderation of their own posts but by broader impacts on self-expression and sociality in their communities. This study also explores the transparency and appeals processes introduced by the EU's 2022 Digital Services Act (DSA) as a primary safeguard against arbitrary and discriminatory moderation. The participants did not see such processes as an adequate response to moderation systems built to enforce a heteronormative "male gaze."

Keywords: social media, content moderation, sexual content moderation, heteronormativity, platform governance

Content moderation—measures to remove or hide content considered harmful or undesirable—is essential to social media (Gillespie, 2018; Roberts, 2019). Yet it has also become a fraught subject of political contestation and, increasingly, regulatory intervention. State censorship and arbitrary private power over online media are longstanding topics of debate. Recent scholarship also increasingly highlights how moderation practices reflect social inequalities, disadvantaging already marginalized groups (Griffin, 2022).

In particular, recent research critiques policies (currently operated by most major Western platforms; Bayley, 2021) that ban sexually explicit content and nudity—typically defined in gendered terms, including "female" but not "male" nipples. Scholarship in cultural studies (Monea, 2022), sociology (Are & Briggs, 2022),

and law (Waldman, 2022) has explored the sexist, transexclusionary and heteronormative constructions of sex, gender, and nudity underlying platforms' policies, suggesting that they disproportionately silence LGBTQIA+ users, as well as other marginalized groups, notably including sex workers.

I build on this existing research with a qualitative investigation of how queer Instagram users in Berlin experience sexual content moderation, making four contributions. First, I add to existing, largely U.S.-based, and relatively synoptic studies on anti-LGBTQIA+ bias in moderation, contributing a detailed account of how it impacts self-expression and social life within a specific European queer community. Second, I suggest that this cannot be reduced to homophobia but involves a broader, interrelated set of repressive norms around gender, sexuality, and the body. Third, I show that these impacts are relational, extending beyond users whose own content is moderated—providing a relevant methodological insight for future empirical work that should consider this broader range of experiences. Finally, I consider the relevance of new regulatory safeguards aiming to address discriminatory moderation in Europe, providing evidence that users may not find them helpful.

Bias and Discrimination in Moderation

Algorithmic bias and discrimination—broad terms describing software systems' tendencies to produce disparate outcomes for different groups, typically reinforcing existing socioeconomic inequalities—are increasingly prominent concerns for researchers and policymakers. These tendencies exist for various reasons, including the use of training datasets, objectives, and categories[2] which reflect structural inequalities, and the fact that technologies are often developed within corporations or other powerful institutions with an interest in maintaining these inequalities (Benjamin, 2019; Miceli, Posada, & Yang, 2022; Noble, 2018). Algorithmic biases are not only pervasive and difficult to mitigate, but are often intractable, or at least not amenable to technical solutions (Balayn & Gürses, 2021; Green, 2022).

Against this backdrop, there is accumulating evidence of widespread (hetero)sexist, racist, classist, and other biases in content moderation. Difficulty accessing platform data has hindered quantitative studies (Leerssen, 2023). However, ethnographic and survey research indicates that many (semi-)professional social media creators perceive discrimination as pervasive (Are & Briggs, 2022; Devito, 2022; Duffy & Meisner, 2022). Qualitative researchers have also shown how moderation policies and software reflect oppressive ideologies, including color-blind racism (Díaz, 2023), heteronormativity (Monea, 2022), and stigmatization of sex work (Are & Briggs, 2022; Blunt & Stardust, 2021), while also systematically overlooking harms predominantly affecting marginalized users, such as hate speech and harassment (Devito, 2022; Griffin, 2022).

In this context, two underlying issues can be conceptually distinguished (although they may be difficult to separate in practice). First, moderation is often biased in the narrow sense that different groups are treated differently for several reasons. Human moderators are generally poorly paid and trained, making decisions under intense time pressure and thus easily influenced by (un)conscious prejudices (Keller & ACLU, 2022; Monea, 2022). Automated moderation tools use datasets of these past biased decisions (Gorwa, Binns, & Katzenbach,

---

[2] For example, categorizing gender as a male/female binary (Keyes, 2018).

2020) and industry-standard datasets known to underrepresent or stereotype minorities (Monea, 2022; Talat, 2021). Unsurprisingly, they exhibit familiar biases, tending to disproportionately associate marginalized identities with negative judgments (Kayser-Bril, 2020; Talat, 2021).

Second, even if they were hypothetically applied completely consistently, the research highlights inequalities built into moderation policies themselves, as well as the tools and heuristics used to operationalize them. For example, filtering keywords often disproportionately silences minorities. because of the use of reclaimed slurs in some marginalized subcultures (Grison & Julliard, 2021; Lux & Lil Miss Hot Mess, 2017). Another well-studied example of inherently discriminatory policies is the removal or algorithmic demotion of content deemed "adult," argued to disadvantage LGBTQIA+ people and sex workers. The following section examines "sexual content moderation" (Waldman, 2022, p. 907) in more detail and introduces the concept of heteronormativity as a lens through which to critique it.

Sexual Content Moderation and Heteronormativity

Are and Briggs (2022) summarized major platforms' approaches as "sex-averse," highlighting a "juxtaposition of sex with harm" (p. 3). In a typical example, Instagram's (n.d.) policy states:

> Instagram is not a place to support or praise terrorism, organized crime or hate groups.
> Offering sexual services, buying or selling firearms, alcohol and tobacco products between
> private individuals, and buying or selling non-medical or pharmaceutical drugs are also
> not allowed. (para. 11)

Sex appears alongside organized crime, hate, and weapons under the heading "follow the law" (Instagram, n.d., para. 11), although in fact many "sexual services" are legal.[3] Instagram also states that "we don't allow nudity," except photos of paintings and sculptures, although these include "some photos of female nipples," with limited exceptions: birth, breastfeeding, "health-related situations," and protest (Instagram, n.d., para. 7). The nudity ban is stated to be "for a variety of reasons," but the only reason explicitly identified is being "appropriate for a diverse audience" (Instagram, n.d., para. 7).

However, scholarship identifies several reasons why most major platforms impose similar policies. These notably include app stores banning pornography (Tiidenberg, 2021), advertisers' preferences for "family-friendly" and "brand-safe" content (Griffin, 2023a), and legal risks. Regarding the latter, the U.S.'s 2018 FOSTA-SESTA legislation—which extended platforms' liability for hosting content related to trafficking and sex work—has been particularly significant (Murray, 2023). In an "American-dominated Internet," it prompted many platforms to introduce stricter policies worldwide, including in countries where sex work is legal (Barwulor, McDonald, Hargittai, & Redmiles, 2021, p. 563; Murray, 2023; Scarlett, 2023). The worldwide imposition of U.S. norms has often attracted criticism from European politicians and media, particularly for suppressing nudity in artistic and political contexts (Griffin, 2023a). However, sexual content bans also reflect local legislation. For example, Germany's youth safety law prohibits adult content on media

---

[3] This obviously depends on jurisdiction, but even jurisdictions banning or severely restricting full-service sex work frequently permit other sexual services like stripping, camming, or online content creation (Scarlett, 2023).

accessible to minors (Barwulor et al., 2021). German media regulators regularly report (otherwise legal) pornographic content to Twitter/X[4] for removal under this law, sometimes also initiating criminal proceedings, and have in particular targeted gay BDSM content (Lulamae, 2023).

The targeting of gay porn in this instance points to broader issues. A growing body of research, often sex worker-led, critiques sexual content moderation as not only excessively restrictive of adult sexual expression (Spišák, Pirjatanniemi, Paalanen, Paasonen, & Vihlman, 2021) but also particularly harmful to sex workers and LGBTQIA+ people (of course not mutually exclusive categories). Sexual content moderation disadvantages these users in both senses identified above. First, they are often targeted for particularly strict and/or arbitrary enforcement, for reasons that include algorithmic bias, biases among human moderators, and reporting by third-party institutions (as in the German example) or other users (Duguay, Burgess, & Suzor, 2020; Monea, 2022; Waldman, 2022). Second, even if consistently applied, these policies encode standards of acceptable sexual expression that are intrinsically exclusionary and heteronormative (Are & Briggs, 2022; Blunt & Stardust, 2021; Coombes, Wolf, Blunt, & Sparks, 2022; Monea, 2022; Waldman, 2022).

At its core, the concept of heteronormativity describes the privileging of heterosexuality and the corresponding marginalization of queer sexualities and cultures (Warner, 1991). However, it also evokes a wider set of interrelated norms around gender, sexuality, and family (Monea, 2022; Robinson, 2016). It is premised on a gender binary that not only erases nonbinary identities but also naturalizes gender inequality through differentiated roles for "masculine men and feminine women" (Robinson, 2016, p. 1) within families and other institutions. Furthermore, heteronormativity does not marginalize all queer people equally. The related concept of "homonormativity" describes the contingent acceptance of queer identities insofar as they assimilate traditional heterosexual norms around family, monogamy, respectability, and economic success—institutions that further exclude LGBTQIA+ people from marginalized racial and class backgrounds (Duggan, 2002).

For Monea (2022) and Waldman (2022), heteronormativity shapes content moderation in several ways. Algorithmic classifiers reproduce biases and stereotypes; user reporting and manual moderation subject minorities to the judgment of heteronormative majority cultures; and since "*sexual* nonnormativity is one defining feature of queerness" (Waldman, 2022, p. 914; emphasis in original), banning public sexual expression inherently suppresses queer identity. Importantly, this does not require explicit homophobia or indiscriminate suppression of all LGBTQIA+-related content (Rodriguez, 2022). Homonormative moderation policies often accept sanitized, "family-friendly" representations of queerness, while coding more sexualized or nonnormative representations as inappropriate (Southerton, Marshall, Aggleton, Rasmussen, & Cover, 2020). For Monea (2022), banning sexual content creates a "digital closet…in which LGBTQIA+ individuals may be allowed to enter the digital public sphere but only so long as they bracket and obscure their sexual identities" (p. 3).

Situated Impacts of Sexual Content Moderation: The Case of Berlin's Queer Community

This literature provides ample cause for concern about the silencing of LGBTQIA+ voices on social media. However, it is primarily U.S.-based, leaving unanswered questions about how moderation impacts LGBTQIA+ communities elsewhere. This research gap is particularly significant, given the growing

---

[4] X (formerly Twitter) is the only major platform that does not generally ban sexual content (Bayley, 2021).

international divergence between legal frameworks regulating moderation. In Europe, for example, the 2022 Digital Services Act (DSA) introduced numerous new regulatory requirements for moderation systems (for a detailed overview, see Wilman, 2023). A key aim is to address fundamental rights issues, including discrimination (see, e.g., Recital 3 DSA). However, since several DSA provisions encourage platforms to intensify efforts to moderate harmful content, commentators have suggested that it could ultimately lead to more overinclusive or arbitrary moderation, rather than less (Barata, 2021; Griffin, 2022; Keller, 2022).

Given these regulatory developments, it is essential to better understand how discriminatory, biased, and unequal moderation plays out in European contexts. Accordingly, this study explores situated experiences of sexual content moderation within a particular European community. Specifically, it explores perceptions and experiences of moderation among LGBTQIA+ Instagram users in Berlin.

German scholars have drawn on U.S.-based research to highlight the risks of anti-LGBTQIA+ discrimination (Dinar, 2021) and have pointed to tensions between Instagram's sexual content policies and European freedom of expression norms (Heldt, 2020). However, aside from journalistic reporting (Lulamae, 2023; Reuter, 2022), to my knowledge, there is only one empirical study of anti-LGBTQIA+ discrimination in content moderation in Germany. Thibeault (2022) documents four U.S.-based and three German-based TikTok creators' perceptions of pervasive anti-LGBTQIA+ bias in moderation and recommendation systems— particularly against people of color and those not conforming to mainstream standards of attractiveness or gender presentation. Thibeault (2022) noted that German interviewees perceived censorship of LGBTQIA+ themes as harsher and more indiscriminate than U.S. interviewees, and stressed the need for further research. Accordingly, this study contributes to the literature with an in-depth study of the experiences of LGBTQIA+ users (not only professional or highly engaged creators) in a specific, geographically bounded community in Germany.

I focused on Berlin partly because of my familiarity with the city[5] and ability to conduct in-person interviews, but also because I considered it a particularly information-rich setting to explore the situated, specific impacts of sexual content moderation. Berlin is known for its vibrant and international queer community, centered in particular around nightlife: Renowned techno clubs coexist and overlap with fetish parties, drag shows, and cruising venues, while also offering spaces for mutual aid, education, and support (Biehl & vom Lehn, 2016; Trott, 2020). Unconventional gender presentation is normalized and often encouraged, as are practices like polyamory, group sex, chemsex, and BDSM. Berlin's queer scene thus offers an ideal setting for empirically exploring Southerton et al.'s (2020) suggestion that sexual content moderation particularly affects more provocative or countercultural representations of queerness.

I focused on Instagram because I observed that it was widely used by young Berliners for diverse purposes, including social life, artistic projects, news, and politics. This allowed me to explore how moderation impacts various types of social media use. Additionally, Instagram was originally developed as a photo-sharing platform, and its technical features, interface, and user cultures retain a strong emphasis on visual content (Leaver, Abidin, & Highfield, 2020). This makes it a particularly relevant site to explore how sexual content moderation affects representations of gender, sexuality, and the body. Finally, compared

---

[5] I do not personally identify as LGBTQIA+, but have spent several years living in Berlin, where I have a largely LGBTQIA+ social circle. My positionality towards the community studied is therefore that of a sympathetic and relatively knowledgeable outsider.

with other widely used platforms, a higher proportion of Instagram users post their own content as well as viewing others' content (Kemp, 2023, p. 188)—meaning typical users could in principle encounter moderation.

These various considerations led me to formulate my first research question:

RQ1:   How are queer Instagram users in Berlin affected by sexual content moderation?

In addition, I aim to shed light on how the DSA might address these impacts. Broadly, the DSA aims to strengthen accountability, transparency, and fairness in content moderation. It pursues this aim primarily via procedural protections, allowing users to understand and contest moderation decisions (Griffin, 2023b; Ortolani, 2022): for example, requiring platforms to publish transparent content policies (Article 14), explain decisions (Article 17), and allow users to appeal (Article 20). I was interested in exploring participants' perceptions of policy transparency and appeals and their relevance to sexual content moderation, prompting my second research question:

RQ2:   How do queer Instagram users in Berlin perceive policies and appeals processes?

Methodology

I investigated these questions through semistructured interviews, a method that generally aims at understanding "social reality" by exploring personal experiences and perspectives (Van Selm & Helberger, 2019, p. 165). Not only is quantitative research limited by restricted access to platform data (Leerssen, 2023); aggregating quantitative data can also obscure minority experiences (Ananny, 2019), and the impacts of moderation depend on contextual factors, including user norms and behavior (Waldman, 2022). Qualitative interviews thus aid in understanding the "social reality" of moderation in particular contexts.

I enlisted acquaintances in Berlin to disseminate a recruitment survey via Instagram. It briefly described the project and requested contact details and some demographic information, with diversity in mind, since research highlights intersectional discrimination in moderation (Duffy & Meisner, 2022; Monea, 2022). Identifying experiences that recur across a diverse sample suggests that they are common within Berlin's queer community. The survey and interviews were conducted in English, since Berlin's young population is highly international (Statistik Berlin Brandenburg, 2023), and in my experience, English rather than German is the primary common language in queer spaces like nightlife venues (although this experience also reflects my position as an immigrant and native English speaker).

Of the 24 respondents who met the eligibility criteria (identifying as queer/LGBTQIA+, using Instagram, and living in Berlin), 12 were interviewed. Table 1 displays demographic information about each participant and their use of Instagram.

Table 1. Interview Participants.

Pseudonym	Gender identity	Sexual orientation	Age range	Nationality	Racial identity	Approx. follower count	Professional, creative, or activist user?
Marina	Cis woman	Queer/ demisexual	22–25	Lithuania	White	550	No
Andy	Genderless (relates to nonbinary and woman)	Attracted to same sex/ gender as themselves	22–25	France	Ashkenazi Jewish	500	Yes (graphic art)
Jennifer	Cis woman	Bisexual/ asexual	31–35	Australia/ Cyprus	Mediterranean	800	Yes (graphic design, psychiatric patient activism)
Pablo	Cis man	Homosexual	31–35	USA	White/Hispanic	570	No
Patrick	Cis man	Asexual	22–25	USA	White	350	In the past (web design)
Fadi	Cis man	Queer	26–30	Lebanon	White	450	No
Stefanie	Cis woman	Pansexual	31–35	Germany	White	1400	Yes (photography)
Ben	Cis man	Queer	31–35	USA	White	1100	Yes (queer activism)
Lina	Butch	Lesbian	22–25	Germany	White	300	No
Marcel	Nonbinary (he/him)	Bisexual	22–25	USA	Biracial (White/Blatino)	1500	Yes (escorting, modelling)
Ava	Nonbinary	Queer/lesbian	26–30	USA	White	200	No
Nathalie	Cis woman	Queer/bisexual	22–25	Germany	White	850	Yes (photography)

Unsurprisingly, since the participants were recruited via Instagram, the sample skewed toward heavy users. All indicated using Instagram "every day/almost every day" or "multiple times a day," although two (Pablo and Andy) periodically made efforts to use it less.

As Table 1 shows,[6] the sample is quite diverse in terms of gender and sexuality, but less so in terms of race and age.[7] Given the survey language and dissemination channels, this represents a sample of one particular queer community—young, predominantly professional, digitally literate, and English-speaking—rather than all LGBTQIA+ people in Berlin. However, since research highlights discrimination against people of color in content moderation (Devito, 2022; Díaz, 2023; Haimson, Delmonaco, Nie, & Wegner, 2021; Salty, 2019), this study's relative racial and class homogeneity is a limitation. Relationships between heteronormativity, class, race, and migration background in content moderation in Europe undoubtedly deserve further empirical research.

My institution's data protection officer and research ethics referent were notified of this research. Following a discussion of the project's aims, interviewees signed and kept a copy of an informed consent form about the recording and use of their data and their ability to withdraw consent. I endeavored to make interviews feel friendly and informal. Participants could choose a place they found comfortable/convenient: Most opted to come to my office (one accompanied by his dog) and one to a café. To facilitate relaxed conversations, I used a topic guide rather than prewritten questions (Arthur & Nazroo, 2014), with three loosely defined sections: First, Instagram's significance in participants' lives and (in their view) for Berlin's queer community, then their experiences and perceptions of moderation,[8] and finally their perspectives on appeals.

I then conducted a thematic analysis, a method commonly combined with qualitative interviews in order to explore subjective experiences and perceptions (Clarke & Braun, 2017; Herzog, Handke, & Hitters, 2019). Thematic analysis aims to identify common patterns across interviews while retaining maximum detail and specificity, aligning with this study's goal of highlighting a particular community's experiences and perspectives. I coded interviews inductively using MAXQDA, following Deterding and Waters' (2018) flexible coding approach. After transcribing and rereading the interviews, I first applied broad codes relating to interview questions and general topics to familiarize myself with the data and compare interviews, before developing more detailed and granular analytic codes. Finally, I compared and connected these codes to construct overarching themes that captured common aspects of the participants' experiences (Herzog et al., 2019).

---

[6] Race, gender, and sexuality are based on free-text questions (Lindqvist, Sendén, & Renström, 2021). I also included a "yes/no" question asking whether participants identify as cisgender, given previous research highlighting trans people face discrimination in moderation (Haimson et al., 2021; Salty, 2019), and phrased it this way to avoid essentializing under one term the various identities and experiences that could fall under "no" (e.g., transgender, agender, nonbinary, etc.)

[7] This could reflect a lack of diversity in the social network disseminating the survey, though I ensured it was shared by people of color. It could also be related to my lacking a budget to pay participants.

[8] I first asked about moderation in very general terms, and only later about sexual content, in order to see which policies participants considered relevant. As discussed later, most brought up sexual content moderation unprompted, reinforcing its significance.

Use of Instagram in Berlin's International Queer Community

Two participants used Instagram for activism (Ben and Jennifer) and five for professional or creative work: graphic art (Jennifer and Andy), photography (Stefanie and Nathalie), and sex work and modelling (Marcel). Nathalie had professional and personal accounts, and Ben helped manage an LGBTQIA+ association's account. All others used only one account.

All participants described Instagram as part of their social lives, providing an easy way to stay in touch with friends, especially those abroad. Several also used it to meet new people, often specifically queer people. Pablo described it as "mainly kind of an extension of Grindr," explaining that after connecting with someone on the dating app, swapping Instagram profiles was a way to share more photos and information. Marcel used it in a similar way with escorting clients. Beyond dating, several participants used Instagram to make friends and socialize in the queer scene, especially those who moved to Berlin without knowing many people:

> I knew like friends of friends, but they're not really in my circles. Most of them weren't queer, so it was kind of just like, I don't know, I wasn't like drawn to that being my core friend group […] I went on Instagram and yeah, I mean, if you sign up for three different clubs or something, and turn on notifications, you immediately are like opened up to everything that's going on. (Ava)

Most participants used Instagram to discover local events, especially queer events like club nights and drag shows. For some, like Ava, this helped them meet people with similar interests. Others used Instagram to share events with friends:

> I usually just post events that I'm interested in and that I think could use some promotion who are not very widely advertised. Mostly queer-related, so like queer electronic scene events or spoken word events or very tiny gigs at bars that no one else would know about. (Marina)

Statements like these suggest Instagram helps connect Berlin's local queer community: It allows people to follow smaller, local, or niche organizations, and exchange events within personal networks, helping maintain a community of like-minded people.

Additionally, many participants mentioned using Instagram to stay informed. Some suggested it offered news that better reflected their political perspectives and/or those of queer communities than mainstream media: For example, Andy said, "Instagram is maybe also beginning to show me like a variety of Jewish communities, and also what it is to be Jewish and queer," while Fadi said, "If you want to see the queer community in Spain, having an issue that came up, I would never read media for such topics. I would rather see them delivering their perspective."

Several others used Instagram to share political messages and opinions—again, often focusing on more personal perspectives than mainstream media. For example, Jennifer shared personal experiences, resources, and critical views on psychiatry, while Andy and Fadi shared information about politics in their

home countries, and Nathalie critiqued the "censoring [and] sexualization of especially female-read bodies" that she encountered when posting nude photography.

Overall, while participants used Instagram in varying ways, these common themes indicate its relevance to queer identity: It enables political and creative self-expression and connections with broader local and international communities. As a channel for political discussions, creative work, and the promotion of offline events, Instagram plays a significant role in queer social life in Berlin. Yet, as the following section shows, participants' accounts suggest that its moderation policies constrain how people within this community can express themselves.

Experiences, Perceptions, and Impacts of Moderation

Only four participants had had their content deleted by Instagram. Two (Nathalie and Marcel) had experienced this multiple times: in both cases, under Instagram's sexual content policies. Both also claimed that their accounts were "shadowbanned," a colloquial term for moderation interventions that reduce the visibility of content, often without informing users (Gillespie, 2022; Leerssen, 2023). Aligning with this research, they explained that it was difficult to search for their profiles, and Nathalie said her professional account had fewer views than her personal account, despite having more followers. Two participants had their content deleted once: Marina had shared nude photos from an art exhibition, and Fadi had created a satirical meme portraying Hezbollah leader Hassan Nasrallah as in a gay relationship with another politician, which was deleted under Instagram's violence and dangerous organizations policy.

Three others described personal experiences involving moderation: Patrick had shared a post that was then deleted as misinformation, Stefanie said she often avoided posting nude photos she thought would be deleted, and Jennifer avoided using drug-related hashtags for activist content, as she believed they led to shadowbanning. Additionally, all 12 participants spoke about their perceptions of Instagram moderation based on how it affected people they know and/or follow. These discussions formed the basis for the first key theme: the relational impacts of moderation. Two further themes were developed: the (hetero)sexist "male gaze" participants saw behind Instagram's policies and its limiting effects on self-expression.

Relational Impacts of Moderation

In the early interviews, it became apparent that even participants with little or no personal experience of moderation were broadly aware of it, sometimes with quite well-informed and critical opinions. When asked about moderation, they often brought up other people's experiences based on conversations on Instagram or offline. For example, Marina said, "Even though I haven't experienced that, a bunch of my friends have, my artist friends, my queer friends."

I started asking in more detail about this, as it seemed to provide important insights. Ultimately, I found that the participants' understandings and experiences of moderation were largely relational. While this finding aligns with prior research on how (semi-)professional users exchange experiences and develop collective understandings of moderation and recommendation systems (Bishop, 2019; Devito, 2022), I stress that even casual users' "algorithmic imaginaries" (Bucher, 2017) are collective and relational. Further,

I argue that not only users' perceptions of moderation but also its concrete impacts should be understood in relational terms. Users are not only concerned by the moderation of their own content but also by its impacts on communities and social networks with which they feel an affinity.

For example, Ben and Patrick—both cis men—expressed concern about unfairness toward women and trans people. Many participants mentioned discrimination toward sex workers and the transexclusionary nature of policies for "female" and "male" nudity, often discussing friends' experiences and taking a position of solidarity with those they considered more affected than themselves. Some participants also described how they were personally affected by the moderation of other people's content:

> This is censorship, but I mean, it's not from the government […] that's kind of scary to see play out and how that kind of contributes to the, like, taboo-ism of being queer or gay lives. (Patrick)

> This is like the thing I'm really sad about. Like I love this body-positive culture as well […] if I don't have it on Instagram, I guess I can definitely find it, like in real life, but it's not the same. (Andy)

Considering these statements and the preceding discussion of how Instagram facilitates queer Berliners' social and cultural lives, it is apparent that the impacts of moderation cannot be fully understood at the individual level. Moderation affects queer users' broader social networks and communities. As such, policies that disproportionately affect certain groups not only disadvantage individuals whose content is targeted but also affect relational and collective interests.

Heteronormative Policies and the Male Gaze

The second theme, the heteronormative male gaze embedded in Instagram's sexual content policies, provides further insights into why they particularly disadvantage LGBTQIA+ communities. While other policy areas (e.g., misinformation, dangerous organizations, and drugs) came up in individual interviews, participants were most aware of sexual content policies (hate speech/racism also came up fairly frequently). When asked about moderation in general terms, eight participants immediately mentioned sexual content. Additionally, several participants who said no when asked if they had had their content removed added, unprompted, that they don't post anything suggestive or obviously queer—with the implication that moderation targets people who do.

An example is this quote from Pablo: "No. [Laughs] Yeah, I think I have one picture on my Instagram that's I'm not wearing a shirt. Yeah, no I haven't." Similarly, Andy responded as follows:

> Not like I posted, like on a post, but also because I don't share…like my art is not queer, like especially.

> *[Interviewer: Your art is not queer?]*

No, it's more like digital, or like nothing that's about sexuality or gender, so…

Participants were generally critical of Instagram's policies, using terms such as "bullshit" (Nathalie), "Puritan bullshit" (Marcel), or "slut-shaming" (Fadi), and expressing fundamental disagreement with interpreting nudity as inherently sexual and sexuality as inherently inappropriate. In Nathalie's words, "Nudity is not porn, nudity is not sexual, nudity is just nudity. So, I think it's bullshit to just censor everything." Marcel expressed similar views:

> There's not even dick in it […] you can see the top of my pubic hair, and they were taken
> down for that. I had some other things similar to this, and then recently I had one that had
> like a nipple in it and they were also like no. What the fuck? What am I going to post here?

Many participants suggested that Instagram's moderation practices essentially reflect the perspectives of cisgender, straight men. Importantly, the gender bias and heterosexism that this implies are understood as closely connected. Stefanie argued that "Nudity that caters the male gaze, like on women, or like heteronormative beauty standards and sexuality like that is accepted […] to me it's a total double standard," while Jennifer said that:

> It just feels like if you're showing too much skin […] or being sexually provocative or
> suggestive, that you're not doing it in this super mainstream, heteronormative way, it
> seems like it's more likely to get flagged.

As these quotes suggest, many participants felt policies were inconsistently applied, with popular commercial accounts and content attractive to a straight "male gaze" treated more leniently:

> So like [Playboy] is a huge account, and they post very explicit stuff where they have only
> the nipples like slightly blurred, like everything else you can see, and that *somehow* is
> okay for Instagram. But then smaller people who are doing sex work are banned because
> they post, I don't know, a bikini photo. Like, really stuff which does not go against
> community guidelines. (Nathalie)

Interestingly, again emphasizing gender bias, Andy (who is nonbinary) felt that cis gay men were treated more leniently than other queer people: "If I compare some of these gay guy memes […] You know they won't get anything. So definitely it's just because like women are supposed to hide their bodies. They'd be like too empowered."

However, as well as criticizing biased enforcement, the participants repeatedly criticized Instagram's underlying policies, singling out two aspects in particular. First, its ban on "female" nipples was repeatedly described as sexist and/or transexclusionary.

> When a person transitions, like what is defined as a woman, and especially with like
> women's nipples. Like what is…just in my mind, it's like so absurd. It's like, it's a nipple!
> I don't understand the need to police and control this. (Ben)

Second, many participants criticized the broader principle of completely banning sex and nudity. Some suggested that this particularly limits queer people's self-expression, as well as other marginalized groups, especially sex workers:

> My own personal perception is that queer people in general […] I don't know, feel more free to express themselves and their bodies, and so express themselves sexually as well, but also physically […] I think the queer community probably posts a lot more of that kind of content than heteronormative people. (Ben)

Many participants linked (hetero)sexism in Instagram's policies to the perspectives of those creating and applying them, presumed to be predominantly straight men. In Andy's evocative description, "I'm not a cis guy, so I think I won't get it exactly, never […] if they don't feel very involved in like this joy, especially queer joy, they will be like, 'Cancel it.'"

Others believed that reporting by (straight) users plays a major role in policy enforcement, suggesting that this also leads to double standards:

> She's wearing a bikini, but she looks very stereotypically attractive, there's going to be a lot of straight men who find that appealing and they like it and they don't report it. But then, if you've got someone who happens to be queer, and maybe they decide they don't want to shave their underarm hair and they have maybe a swimsuit on that doesn't look stereotypically attractive, you know, it's more likely that a bunch of conservative dudes are going to report that. (Jennifer)

This echoes Waldman's (2022) suggestion that since queer sexuality is definitionally nonnormative, reporting-based moderation systems that reflect majority audiences' judgments of appropriateness will inherently tend to suppress queer content. Importantly, however, the participants' accounts suggest that these impacts on queer communities cannot simply be ascribed to homophobia. Rather, they perceived moderation systems as enforcing a broader, interrelated set of norms around (un)acceptable depictions of gender, sexuality, and bodies, ultimately dominated by cis-straight male perspectives and excluding people whose self-presentation offends this "male gaze."

Limiting Queer Self-Expression

The final theme captures the concrete impact of these heteronormative systems. While the participants' personal experiences of moderation varied significantly, a common theme was that Instagram's sexual content policies limited queer people's personal and creative self-expression. Stefanie and Nathalie, who both did nude photography, found them particularly limiting and frustrating: for example, saying, "It's just annoying [laughs] […] I always have to censor things and like think about what can I post, what can I not" (Nathalie).

However, as emphasized in the discussion of relationality, these constraints not only affect artists but also broader communities of users interested in such content. For example, Andy expressed frustration

with Instagram's limitations on body-positivity content, while Marina could not share photos of an exhibition she wanted to recommend:

> It was feminist, it was queer, it was groundbreaking. It was really, really well done and I really enjoyed it. And Instagram took down most of my posts that showed nudity, which was really sad because the pictures were not even overtly sexual. (Marina)

First- and second-hand experiences of moderation also affected how participants used Instagram. A majority had avoided posting content, noticed others doing it, or presented it differently, effectively self-censoring. In addition to simply not posting content or using certain hashtags, participants frequently mentioned misspelling words in captions to circumvent keyword filtering, and concealing body parts in images with emojis or blurs. Previous literature has explored misspelling and hashtag avoidance as tactics to resist moderation practices that users perceive as suppressing LGBTQIA+ content and important topics such as mental health (Klug, Steen, & Yurechko, 2023; Thibeault, 2022). However, by making content less recognizable for moderation algorithms, they also make it less accessible to intended audiences (Glatt & Banet-Weiser, 2021), especially users relying on screen-readers (Coombes et al., 2022).

As such, self-censorship represents not only another constraint on queer users' self-expression but also another way that moderation indirectly impacts others. In the context of Instagram—a visual platform widely used by artists—the finding that nudity policies push users to censor photos through additions like emojis is also worth highlighting as a constraint on artistic expression:

> It's really aesthetic and I really would have liked to put it on Instagram, but I was like, they're going to take it down anyway, I think, and I don't want to like censor it […] I don't want to censor something that should not be censored, I felt like, putting emojis on the boobs or whatever, you know. (Stefanie)

Finally, many participants felt that policies were applied inconsistently and arbitrarily, so efforts to comply were no guarantee against content or account removal. No participants had personally experienced the latter, but some brought it up as a particularly consequential risk for artists, sex workers, and other professional users. For example, in referring to a friend, Stefanie said, "That's her whole business, and she was really careful about it, like to know the guidelines and follow everything […] she had to start from zero again."

Stefanie and Nathalie described Instagram as effectively mandatory for photographers, while Marcel suggested that for modelling, "it has become quite essential," as other channels are expensive and less accessible. Moderation can therefore have serious material impacts, including loss of income and professional resources (Are & Briggs, 2022).

Overall, then, participants not only personally disagreed with Instagram's gendered, heteronormative policies—they saw consequential impacts on their own and others' freedom to express themselves in ways involving nudity, sexuality, and the body. Evidently, this does not only affect queer people. However, statements like Stefanie's refusal to "censor something that should not be censored"

or Ben's questioning of the need to "police and control" nipples suggest that in Berlin's queer community, there is a widespread expectation that people should be allowed to share and engage with content involving nudity and sex. Instagram's policies thus significantly limit how this community can express itself and interact online.

Experiences and Perceptions of Appeals

Finally, since policy transparency and appeals feature prominently in the incipient EU regulatory framework and are intended to address issues such as those discussed above, I asked participants about their perceptions of Instagram's existing transparency and appeals processes and their relevance. Two key themes emerged. First, existing published policies were not considered particularly relevant or useful. Second, although a few participants had personally used appeals, those who were aware of them largely perceived them as arbitrary and unhelpful.

Opaque and Irrelevant Guidelines

Most participants described Instagram's rules on sexual content and/or hate speech in broad terms, but—except Nathalie, who described Instagram's nudity policies and their exceptions in detail— none had read or had detailed familiarity with its published guidelines. Generally, this was considered an unrealistic expectation, at least for nonprofessional users: "Who's going to read that, for one? When you accept the agreement, no one really reads that, which is true, it's our responsibility, but it's hidden in there." (Marcel)

Many participants suggested that reading the guidelines would not be useful, as they are not consistently followed: For example, Marcel added, "I mean, I think they say one thing and do another anyway." Moderation was generally considered arbitrary and selective rather than following established policies. Nathalie said, "Sometimes I post a post which stays as a post to the story, but then the story gets deleted, which doesn't make any sense […] I don't see a pattern." Fadi suggested, "If enough people report it, regardless of what the content is, it's going to be removed. Because like I don't see any proper consistency at all."

This suggests that for these participants, requiring platforms to publish more transparent policies would not be considered particularly helpful, since there is little trust that they will accurately state or consistently follow these policies.

Appeals as Arbitrary and Unhelpful

Only three participants—Nathalie, Fadi, and Marcel—had appealed moderation decisions. Although Nathalie said she did so regularly and found it straightforward, none of them felt that decisions were carefully reviewed or that consistent rules were applied. Participants had not received useful explanations, either with the initial decision or on appeal. Fadi said, "They just repeated what they said the first time, so I didn't understand," and shared a screenshot (see Figure 1) showing that the rejected appeal simply cited "the violent and dangerous organizations policy" with no further explanation:
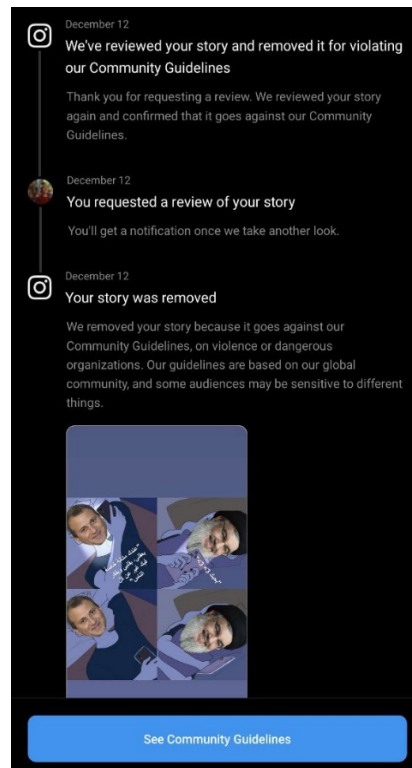
Figure 1. Screenshot provided by "Fadi."

As with moderation generally, outcomes were described as arbitrary or "kind of random" (Nathalie) and as box-ticking rather than meaningful reconsideration. Participants generally expressed little trust in Instagram and its moderators, suggesting that decisions could be influenced by prejudice or commercial incentives. In Jennifer's words, for example, "I'm very skeptical about the moves social media leaders make in general. I don't know if they necessarily want the best for their users. I feel like what they want is the best thing for their advertisers."

Participants not only felt policies were wrongly or inconsistently applied but disagreed fundamentally with their heteronormative approach to nudity and sexuality. Importantly, existing appeals processes, and those mandated by the DSA, do not address this; they allow contestation of individual decisions but not underlying policies (Griffin, 2023b). Marina suggested that the priority should instead be more inclusive policy development:

> I think from the ground up they need to reassess the rules that they already have in place, instead of make it possible for us to complain and for them to be like, "OK, tick, I responded to this complaint. We're all sorted." […] There need to be people of more diverse backgrounds who are investigating their policy.

This small sample does not provide enough data to draw significant conclusions about how Instagram's appeals work. However, these accounts point to the limited usefulness of appeals processes that do not provide meaningful information and opportunities to challenge decisions, and they suggest that users find such processes unhelpful. This tracks other recent research involving larger samples (Are, 2023). As the DSA institutionalizes and extends these procedures, further research into how they function in practice is essential.

Conclusion

Building on U.S. literature critiquing heteronormative moderation practices (Monea, 2022; Waldman, 2022), this study has presented qualitative evidence of their tangible impacts on queer art, sociality, and self-expression in a European context. Documenting frustration with Instagram's "Puritan bullshit" policies within a specific community that normalizes and values unconventional gender and sexual expression lends support to Southerton et al.'s (2020) argument that platforms' "homonormative" policies have particular impacts on more unconventional and sexualized representations of queer identity. Participants perceived Instagram as enforcing repressive, (hetero)sexist, and trans- and sex-worker-exclusionary norms around nudity, gender, and sexuality. They understood these policies as heteronormative in the broadest sense—denoting not only homophobia but also a "male gaze" that privileges content deemed appealing by a mainstream culture dominated by straight cisgender men, while policing and controlling women's bodies, minority gender identities, and unconventional sexual and artistic expression.

My findings additionally highlight these policies' relational and collective impacts: They should not just be understood as impacting specific users who share such content, but as having ripple effects on their social networks. Ultimately, this heteronormative policing of media representations of gender and sexuality also has broader social implications. For example, research highlights the importance for young LGBTQIA+ people of access to information about diverse queer identities and experiences, not only those that fit the heteronormative and exclusionary standards of "brand safety" and "family-friendliness" (Devito, 2022; Waldman, 2022).

Finally, by exploring users' perceptions of policy transparency and appeals, this study sheds light on the approach to regulating moderation now institutionalized in the DSA. First, it casts doubt on the practical usefulness of policy transparency and appeals. Echoing other recent research on "dysfunctional appeals" (Are, 2023), participants generally felt that stated policies misrepresented actual moderation practices and perceived enforcement as arbitrary and biased, with appeals essentially a box-ticking exercise. Second, many participants perceived moderation as fundamentally shaped by cis straight male perspectives, prejudices, and commercial priorities. As such, even more effective and meaningful transparency and appeals would not address the underlying problem—that Instagram's policies and the systems set up to implement them are designed to enforce exclusionary and repressive norms around sexuality and gender. This suggests reasons for skepticism toward the DSA's procedurally focused approach and points to the need for a more fundamental reconsideration of how moderation systems are designed and whose interests and ideologies they serve.

References

Ananny, M. (2019, August 21). *Probably speech, maybe free: Toward a probabilistic understanding of online expression and platform governance*. Knight Foundation. Retrieved from https://knightcolumbia.org/content/probably-speech-maybe-free-toward-a-probabilistic-understanding-of-online-expression-and-platform-governance

Are, C. (2023, April 3–4). *"Dysfunctional" appeals and failures of algorithmic justice in Instagram and TikTok content moderation*. Paper presented at PlatGov Research Network Conference (virtual).

Are, C., & Briggs, P. (2022). The emotional and financial impact of de-platforming on creators at the margins. *Social Media + Society, 9*(1), 1–12. doi:10.1177/20563051231155103

Arthur, S., & Nazroo, J. (2014). Designing fieldwork strategies and materials. In J. Ritchie & J. Lewis (Eds.), *Qualitative research practice: A guide for social science students & researchers* (pp. 109–137). Thousand Oaks, CA: SAGE.

Balayn, A., & Gürses, S. (2021, September 21). *Beyond debiasing: Regulating AI and its inequalities*. Retrieved from https://edri.org/our-work/if-ai-is-the-problem-is-debiasing-the-solution/

Barata, J. (2021, July 27). *The Digital Services Act and its impact on the right to freedom of expression: Special focus on risk mitigation obligations*. DSA Observatory. Retrieved from https://dsa-observatory.eu/2021/07/27/the-digital-services-act-and-its-impact-on-the-right-to-freedom-of-expression-special-focus-on-risk-mitigation-obligations/

Barwulor, C., McDonald, A., Hargittai, E., & Redmiles, E. (2021). 'Disadvantaged in the American-dominated Internet': Sex, work, and technology. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, 563*, 1–16. doi:10.1145/3411764.3445378

Bayley, C. (2021, May 25). Sexual censorship on social media: What I learned. *Perspectives on Public Purpose for Emerging Technologies*. Retrieved from https://www.belfercenter.org/publication/sexual-censorship-social-media-what-i-learned

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the New Jim Code*. Cambridge, MA: Polity.

Biehl, B., & Von Lehm, D. (2016). Four-to-the-floor: The techno discourse and aesthetic work in Berlin. *Society, 53*, 608–613. Retrieved from https://link.springer.com/article/10.1007/s12115-016-0083-8

Bishop, S. (2019). Managing visibility on YouTube through algorithmic gossip. *New Media & Society, 21*(11–12), 2589–2606. doi:10.1177/1461444819854731

Blunt, D., & Stardust, Z. (2021). Automating whorephobia: Sex, technology and the violence of deplatforming: An interview with Hacking//Hustling. *Porn Studies, 8*(4), 350–366. doi:10.1080/23268743.2021.1947883

Bucher, T. (2017). The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society, 20*(1), 30–44. doi:10.1080/1369118X.2016.1154086

Clarke, V., & Braun, V. (2017). Thematic analysis. *Journal of Positive Psychology, 12*(3), 297–298. doi:10.1080/17439760.2016.1262613

Coombes, E., Wolf, A., Blunt, D., & Sparks, K. (2022). Disabled sex workers' fight for digital rights, platform accessibility, and design justice. *Disability Studies Quarterly, 42*(2). doi:10.18061/dsq.v42i2.9097

Deterding, N. M., & Waters, M. C. (2018). Flexible coding of in-depth interviews: A twenty-first-century approach. *Sociological Methods & Research, 50*(2), 708–739. doi:10.1177/0049124118799377

Devito, M. A. (2022). How transfeminine TikTok creators navigate the algorithmic trap of visibility via folk theorization. *Proceedings of the ACM on Human-Computer Interaction, 6*(CSCW2), Article No.: 380, 1–31. doi:10.1145/3555105

Díaz, Á. (2023). Online racialization and the myth of colorblind content policy. *Boston University Law Review, 103*. Advance online publication. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4509790

Dinar, C. (2021). *The state of content moderation for the LGBTIQA+ community and the role of the EU Digital Services Act*. Heinrich-Böll-Stiftung. Retrieved from https://us.boell.org/sites/default/files/importedFiles/2021/06/21/HBS-e-paper-state-platform-moderation-for-LGBTQI-200621_FINAL.pdf

Duffy, B. E., & Meisner, C. (2022). Platform governance at the margins: Social media creators' experiences with algorithmic (in)visibility. *Media, Culture & Society, 45*(2), 285–304. doi:10.1177/01634437221111923

Duggan, L. (2002). The new homonormativity: The sexual politics of neoliberalism. In R. Castronovo & D. D. Nelson (Eds.), *Materializing democracy: Toward a revitalized cultural politics* (pp. 175–194). Durham, NC: Duke University Press.

Duguay, S., Burgess, J., & Suzor, N. (2020). Queer women's experiences of patchwork platform governance on Tinder, Instagram, and Vine. *Convergence, 26*(2), 237–252. doi:10.1177/1354856518781530

Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. New Haven, CT: Yale University Press.

Gillespie, T. (2022). Do not recommend? Reduction as a form of content moderation. *Social Media + Society, 8*(3), 1–13. doi:10.1177/20563051221117552

Glatt, Z., & Banet-Weiser, S. (2021). Productive ambivalence, economies of visibility and the political potential of feminist YouTubers. In S. Cunningham & D. Craig (Eds.), *Creator culture: An introduction to global social media entertainment* (pp. 39–56). Berlin, Germany: De Gruyter.

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society, 7*(1), 1–15. doi:10.1177/2053951719897945

Green, B. (2022). Escaping the impossibility of fairness: From formal to substantive algorithmic fairness. *Philosophy & Technology, 35*(4), 90, 1–32. doi:10.1007/s13347-022-00584-6

Griffin, R. (2022). The sanitised platform. *JIPITEC, 13*(1), 36–52. Retrieved from https://www.jipitec.eu/issues/jipitec-13-1-2022/

Griffin, R. (2023a). From brand safety to suitability: Advertisers in platform governance. *Internet Policy Review, 12*(3), 1–25. doi:10.14763/2023.3.1716

Griffin, R. (2023b). Rethinking rights in social media governance: Human rights, ideology and inequality. *European Law Open, 2*(1), 30–56. doi:10.1017/elo.2023.7

Grison, T., & Julliard, V. (2021). Les enjeux de la modération automatisée sur les réseaux sociaux numériques: Les mobilisations LGBT contre la loi Avia [The challenges of automated moderation on digital social media: LGBT mobilisations against the Avia law]. *Communication, Technologies et Développement, 10*, 1–10. doi:10.4000/ctd.6049

Haimson, O. L., Delmonaco, D., Nie, P., & Wegner, A. (2021). Disproportionate removals and differing content moderation experiences for conservative, transgender, and Black social media users: Marginalization and moderation gray areas*. Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW2), 466. doi:10.1145/3479610

Heldt, A. (2020, October 15). Borderline speech: Caught in a free speech limbo? *Internet Policy Review: Opinion*. Retrieved from https://policyreview.info/articles/news/borderline-speech-caught-free-speech-limbo/1510

Herzog, C., Handke, C., & Hitters, E. (2019). Analyzing talk and text II: Thematic analysis. In H. Van den Bulck, M. Puppis, K. Donders, & L. Van Audenhove (Eds.), *The Palgrave handbook of methods for media policy research* (pp. 385–401). London, UK: Palgrave MacMillan. Retrieved from https://link.springer.com/chapter/10.1007/978-3-030-16065-4_22

Instagram. (n.d.). *Community guidelines*. Instagram help centre. Retrieved September 6 2023, from https://help.instagram.com/477434105621119/?helpref=uf_share

Kayser-Bril, N. (2020, May 19). *Automated moderation tool from Google rates People of Color and gays as "toxic."* AlgorithmWatch. Retrieved from https://algorithmwatch.org/en/automated-moderation-perspective-bias/

Keller, D. (2022, November 7). *The EU's new Digital Services Act and the rest of the world*. Verfassungsblog. Retrieved from https://verfassungsblog.de/dsa-rest-of-world/

Keller, D., & ACLU. (2022, August 23). *Daphne Keller and ACLU file comment to Meta Oversight Board in "UK Drill Music" case*. Stanford Cyber Policy Center. Retrieved from https://cyber.fsi.stanford.edu/news/daphne-keller-and-aclu-file-comment-uk-drill-music-case

Kemp, S. (2023). *Digital 2023: Global overview report.* Datareportal. Retrieved from https://datareportal.com/reports/digital-2023-global-overview-report

Keyes, O. (2018). The misgendering machines: Trans/HCI implications of automatic gender recognition. *Proceedings of the ACM on Human-Computer Interaction, 2*(CSCW), 1–22. doi:10.1145/3274357

Klug, D., Steen, E., & Yurechko, K. (2023). How algorithm awareness impacts Algospeak use on TikTok. *WWW '23 Companion: Companion Proceedings of the ACM Web Conference 2023*, 234–237. doi:10.1145/3543873.3587355

Leaver, T., Abidin, C., & Highfield, T. (2020). *Instagram: Visual social media cultures*. Hoboken, NJ: Wiley.

Leerssen, P. (2023). *Seeing what others are seeing: Studies in the regulation of transparency for social media recommender systems* (PhD thesis). University of Amsterdam, Amsterdam, Netherlands. Retrieved from https://hdl.handle.net/11245.1/18c6e9a0-1530-4e70-b9a6-35fb37873d13

Lindqvist, A., Sendén, M. G., & Renström, E. A. (2021). What is gender, anyway: A review of the options for operationalising gender. *Psychology & Sexuality, 12*(4), 332–344. doi:10.1080/19419899.2020.1729844

Lulamae, J. (2023, April 4). *AI surveillance rumors: Gay adult content creators face sanctions*. AlgorithmWatch. Retrieved from https://algorithmwatch.org/en/gay-adult-content-sanctions/

Lux, D., & Lil Miss Hot Mess. (2017, August 14). Facebook's hate speech policies censor marginalized users. *Wired*. Retrieved from https://www.wired.com/story/facebooks-hate-speech-policies-censor-marginalized-users/

Miceli, M., Posada, J., & Yang, T. (2022). Studying up machine learning data: Why talk about bias when we mean power? *Proceedings of the ACM on Human-Computer Interaction, 6*, 1–14. doi:10.1145/3492853

Monea, A. (2022). *The digital closet: How the Internet became straight*. Cambridge, MA: MIT Press. doi:10.7551/mitpress/12551.001.0001

Murray, Z. (2023, June 30). *How US law FOSTA-SESTA "inadvertently" censored sex workers' social media usage*. Sciences Po Digital, Governance & Sovereignty Chair. Retrieved from https://www.sciencespo.fr/public/chaire-numerique/en/2023/06/30/student-essay-how-us-law-fosta-sesta-inadvertently-censored-sex-workers-social-media-usage/

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York: NYU Press.

Ortolani, P. (2022, November 7). *If you build it, they will come*. Verfassungsblog. Retrieved from https://verfassungsblog.de/dsa-build-it/

Reuter, M. (2022, October 5). TikTok würgt Diskussion über LGBTQ, Heterosexuelle und Drogen ab [TikTok suppresses discussion of LGBTQ, straight people and drugs]. *Netzpolitik*. Retrieved from https://netzpolitik.org/2022/informationskontrolle-tiktok-wuergt-diskussion-ueber-lgbtq-heterosexuelle-und-drogen-ab/

Roberts, S. (2019). *Behind the screen: Content moderation in the shadows of social media*. New Haven, CT: Yale University Press.

Robinson, B. A. (2016). Heteronormativity and homonormativity. In A. Wong, M. Wickramasinghe, R. Hoogland, & N. A. Naples (Eds.), *The Wiley Blackwell encyclopedia of gender and sexuality studies*. Hoboken, NJ: Wiley-Blackwell. doi:10.1002/9781118663219.wbegss013

Rodriguez, J. A. (2022). LGBTQ incorporated: YouTube and the management of diversity. *Journal of Homosexuality, 70*(9), 1807–1828. doi:10.1080/00918369.2022.2042664

Salty. (2019). *An investigation into algorithmic bias in content policing on Instagram*. Retrieved from https://saltyworld.net/algorithmicbiasreport-2/

Scarlett, M. (2023, April 12). Five years of FOSTA-SESTA. *Red Pepper*. Retrieved from https://www.redpepper.org.uk/fosta-sesta-sex-worker-exploitation/

Southerton, C., Marshall, D., Aggleton, P., Rasmussen, M. L., & Cover, R. (2020). Restricted modes: Social media, content classification and LGBTQ sexual citizenship. *New Media & Society, 23*(5), 920–938. doi:10.1177/1461444820904362

Spišák, S., Pirjatanniemi, E., Paalanen, T., Paasonen, S., & Vihlman, M. (2021). Social networking sites' gag order: Commercial content moderation's adverse implications for fundamental sexual rights and wellbeing. *Social Media & Society, 7*(2), 1–9. doi:10.1177/20563051211024962

Statistik Berlin Brandenburg. (2023, February 14). *Einwohnerregisterstatistik 31.12.2022 Berlin* [Population register statistics 31.12.2022 Berlin]. Statistik Berlin Brandenburg. Retrieved from https://www.statistik-berlin-brandenburg.de/029-2023

Talat, Z. (2021) *"It ain't all good": Machinic abuse detection and marginalization in machine learning* (PhD thesis). University of Sheffield, Sheffield, UK. Retrieved from https://etheses.whiterose.ac.uk/30950/

Thibeault, S. (2022). *Die Zensur von LGBTQ+-Themen auf TikTok in Deutschland und den USA: wie Content-Creator:innen ihr entgegenwirken* [Censorship of LGBTQ+ themes on TikTok in Germany and the US: How content creators resist them] (MA thesis). Technische Hochschule Köln, Cologne, Germany. Retrieved from https://publiscologne.th-koeln.de/frontdoor/deliver/index/docId/2084/file/MA_Thibeault_Sarah.pdf

Tiidenberg, K. (2021). Sex, power and platform governance. *Porn Studies, 8*(4), 381–393. doi:10.1080/23268743.2021.1974312

Trott, B. (2020). Queer Berlin and the Covid-19 crisis: A politics of contact and ethics of care. *Interface, 12*(1), 88–108. Retrieved from https://www.interfacejournal.net/wp-content/uploads/2020/07/Interface-12-1-Trott.pdf

Van Selm, M., & Helberger, N. (2019). Talking to people II: Qualitative interviews. In H. Van den Bulck, M. Puppis, K. Donders, & L. Van Audenhove (Eds.), *The Palgrave handbook of methods for media policy research* (pp. 161–177). London, UK: Palgrave MacMillan. Retrieved from https://link.springer.com/chapter/10.1007/978-3-030-16065-4_9

Waldman, A. E. (2022). Disorderly content. *Washington Law Review, 97*(4), 907–976. Retrieved from https://digitalcommons.law.uw.edu/wlr/vol97/iss4/8/

Warner, M. (1991). Introduction: Fear of a queer planet. *Social Text, 29*, 3–17. Retrieved from https://www.jstor.org/stable/466295

Wilman, F. (2023). *The Digital Services Act (DSA)—An overview*. SSRN. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4304586