Thibault Grison

# Playing Hide and Seek with Algorithms in the "Gay TikTok": From Shadowbanning to Platform Affordances

**Abstract:** This chapter investigates algorithmic bypass strategies that sexual and gendered communities set up on TikTok to post sexual content on the platform. Focusing on the gay community, the chapter proposes a typology of platform affordances through sound, text and image. The analysis draws on a digital ethnography method consisting of "growing a niche algorithm" to study content recommendation and invisibilisation on social media. Based on an epistemology of the closet, the chapter concludes with a discussion on online visibility and virality as matters of adjustment and negotiation.

**Keywords:** platform affordances, TikTok, LGBT, sex, content moderation

A great deal of queer and feminist research has pointed out how difficult it is for sexual and gendered communities to make themselves visible both offline and online. While the use of social networks to build one's sexual and gender identity and to exchange around these issues has been widely documented – through objects of study such as online activism, for example (see Jouët 2022; Despontin Lefevre 2022; Armangau and Figeac 2023, etc.), other work is also emerging on the technical conditions of their invisibilisation (see Monea 2022; Gillespie 2018). This chapter focuses on the practices put in place by LGBT communities to avoid censorship of their sexual content on TikTok. Inspired by approaches relating to the epistemology of the closet (see Sedgwick 1990), I'm interested in the misappropriated uses of the application's functionalities by queer internet users. In other words, I study how platform affordances imagined by subaltern communities (see Tiidenberg and van der Nagel 2020) shed light on the opaque workings of platform moderation *dispositifs*.

The virality of content posted online is primarily conditioned by the moderation, censorship and recommendation rules and mechanisms that govern platforms (see Gillespie et al. 2020). These systems and rules evolve over time (see Chan, Su and Shore 2023) vary from one language and country to another, and above all, diverge from one social networking site (SNS) to another (see Badouard 2021). This moderation work manifests itself in different forms, ranging from the deletion of a post or content to its dereferencing on recommendation feeds. For example, since


https://doi.org/10.1515/9783111311371-013

around 2019, and on TikTok in particular, many internet users and journalists have been denouncing "shadowbanning" phenomena. Shadowbanning is a form of moderation that consists of making content or a user profile invisible, without the creator of the content being aware of it or being informed of it, meaning that this content has been de-referenced from users' feeds (see Gillespie 2018). I would argue that shadowbanning can be considered a form of "insidious censorship" (Fergus et al. 2020) because it allows companies to moderate content without being seen doing it. This form of censorship impacts the circulation of content by making it invisible, without making it disappear. Due to TikTok's relatively strict moderation policies prior to 2021 regarding so-called "political", "sexual" or "violent" content, various activist communities widely denounced the censorship they were experiencing. In France, this movement took off during the LGBT mobilisations against the French AVIA law and led to the circulation of a range of algorithmic circumvention strategies or denunciation campaigns against censorship (see Grison and Julliard, 2021). While these practices are primarily aimed at spreading content from LGBT communities across the platform, the issue of visibility is also subject to self-moderation tactics (see Seering 2020) to fly under the radar of algorithmic tracking or to avoid waves of homophobic harassment. As too much visibility could lead to their account being banned or to a violent raid, internet users play a form of hide-and-seek with algorithmic moderation processes, switching between a desire for visibility and self-preservation.

The chapter is therefore based less on a general analysis of shadowbanning and the discourses about it that circulate online, but rather on the specific tactics implemented by individuals who experience discrimination and silencing dynamics in their everyday life both in and offline. These practices are to be considered as the *savoir-faire* of epistemic and situated communities (see Dell'Omodarme 2014) or "savoirs-d'expérience", depending on the context. In this essay, I analysed them using techno-semiotical approaches and in regard with the notion of "platform affordance" as defined by Tiidenberg and van der Nagel (2020)[1]. My argument is that by starting from these queer tactics of circumventing algorithmic censorship, social sciences scholars may approach content moderation beyond its opacity, but rather as a process of distribution of negotiated regimes of visuality.

This chapter is based on a digital ethnography investigation initiated in February 2021 on TikTok. I draw on the last part of a corpus I collected on the app since the beginning of my PhD thesis, which investigates the impact of content

1 As ways in which features are twisted, circumvented and re-appropriated by internet users.

moderation policies on sexual and gendered communities. This corpus features around 100 videos collected since 2021 and related to online gay sexuality[2].

First, I look back at the links between sex and content moderation; then I outline the method of collecting and analysing the video content I work on, being careful to explain the methodological challenges involved when investigating the invisibilisation of online content; finally, presenting a typology of platform affordances, I discuss how these tactics turn into collective practices designed to make sexual and gendered communities' content strategically (in)visible online.

# 1 Sex, Content Moderation, Algorithms and Social Media

## 1.1 "Porn made the Internet"

Porn studies have played a major role in highlighting the responsibility of digital devices in crystallising and reproducing gender stereotypes online. The pornography industry, for example, plays a part in the eroticisation and naturalisation of specific social relations, the invisibilisation of gender violence and the reinforcement of a large number of gender, race and class stereotypes that structure our society (see Jahjah 2022; Benjamin 2019). Thus, pornography is still today a privileged object of study for the construction of sexual scripts (see Gagnon 1999) and sexual and gendered identities (see Damian-Gaillard 2014). It is therefore interesting to observe how porn studies researchers have been led to seize on the study of algorithmic technologies to enrich their work. Computer vision-based pornography filtering algorithms (CVPF) are thought to be primarily responsible for crystallising discriminatory representations of sexual and gender relations online. A critical review of the scientific literature (see Robert Gehl, Lucas Moyer-Horner and Sara Yeo 2017) on these technologies shows how the implementation of cisgender men's cognitive biases is indeed at the heart of the design of these machines insofar as the design teams are essentially made up of white cisgender heterosexual men (see Noble 2018; Jean 2019). The result is a heterocentric and phallocentric vision of pornography and therefore of the internet. These technologies are also widely used for the ("good") functioning of other recommendation

2 By sexual content, we mean any content that refers in any way to sexuality, including pornography, erotic images, nudity, sexual narratives, sex education, political content and so on. On the difficulty of defining sexuality online, see Tiidenberg and van der Nagel (2020); Paasonen, Jarrett and Light (2020).

systems (news feeds, trending topics, etc.), whatever the field. These reuses go hand in hand with the implementation of biases and can therefore have a significant effect on freedom of speech and online visibility (see Are 2020).

In fact, the gradual prohibition of sexually explicit content from SNSs – enabled by the rollout of such technologies – also has consequences for sexual and gendered communities, in particular sex workers who rely on these platforms to carry out their business, generate income or verify the identity of, and engage with, potential clients (see Tiidenberg and van der Nagel 2020). In fact, since 2018 with the banning of content classified as NSFW by Tumblr, the status of sexuality on SNSs has become a crucial stake for these platforms: "This use of the NSFW category participates in a definition of nudity, sex and sexuality as problematic realities to be excluded from social media" (Paasonen et al. 2020, 167).

## 1.2 The Risk of Censorship for Sexual and Gendered Communities on SNSs

One of the main sources of algorithmic bias towards LGBT people online lies in the assimilation of sexual identity with stereotypes related to sexual practices and preferences. In the context of online content moderation, Tarleton Gillespie (2018) explains, in his chapter "To remove or to filter", how Tumblr's recommendation algorithms blocked "#gay" search results to combat the proliferation of pornographic content. The moderation logic was as follows: Insofar as a lot of sexually explicit content is indexed under the term "gay", this term is made invisible to inhibit the amount of pornographic content on the platform. This (de)referencing then leads to the censoring of non-pornographic gay content posted by homosexual users, who are *de facto* automatically assimilated to sexually explicit content (see Grison and Julliard 2021).

While the LGBT community is often assimilated to sex and marked by stereotyped sexual practices, it's worth remembering that LGBT communities have also structured their activism around sexual liberation discourses. Beyond activism, sexual culture is very present in discussions and interactions within gay communities (see Vörös 2020). The result is a "queer world" impossible to map or restrict to delimited communities and spaces other than through the recognition of practices, discourses and investments of self in subaltern sexual affects and cultures (see Berlant and Warner 2018). SNSs are also spaces in which we exchange and interact according to a semio-discursive repertoire in which the question of sex is everywhere present (see Berlant and Warner 2018). These can range from sharing experiences to encoded humouristic content or a whole set of explicit sexual visuals. In particular, SNSs like Tumblr (until 2018) or Twitter are heavily invested in

by so-called subaltern communities to consume amateur "for adults" content (see Cao 2021). On TikTok, such content would be considered illegal and would be systematically moderated[3]. However, since an update in March 2023, TikTok has clarified its community rules on sexuality, explaining that sex education content is authorised, and that body exposure with little clothing is tolerated but will not appear on users' For You Page (FYP)[4]. Finally, it should be noted that TikTok is used massively by homosexual sex workers to redirect users to dedicated platforms such as OnlyFans to view sexually explicit content, or even Twitter, whose moderation rules are less strict (see van der Nagel 2021). The result is an initial discrepancy between what is *a priori* visible and permitted on the platform and what is actually published and seen by others.

## 1.3 Being Gay Online: A New "Digital Closet"

The epistemology of the closet (Sedgwick 1990) examines power systems, social norms and individual consequences associated with the concealment of one's queer identity. It is a conceptual framework widely used in queer studies to analyse both the conditions of assignment of these communities and their agency potential. Unlike vernacular discourses about "coming out [of the closet]" stories, queer theory considers that the closet is made of a set of practices to adjust to what extent one queer individual can live as gay, lesbian, trans, bi, queer, etc. in public spaces. Thus, considering social media as a "digital closet" (see Monea 2022) implies having a closer look at how LGBT individuals "are forced to digitally segregate that aspect of themselves from their everyday online existence. To not have your account banned, to not have your content censored, to not find yourself demonetized, or, in short, to participate in this new internet-mediated world of ours, you must relegate a certain part of your identity to a digital closet" (Monea 2022, 181). The quest for virality is not always desired and desirable regarding the risk of exposure – synonymous with being harassed, reported, or censored – for LGBT individuals and activists online. In his article "Outed by the Machine", Alexander Cho (2017) explains how the design of social media like Facebook does not necessarily allow queer internet users to live their sexual and gender identity as they see fit. To put it differently, marginalised communities overall are caught in a vice between forced overexposure and invisibility of their voices and representations. Therefore, rely-

**3** https://www.tiktok.com/community-guidelines/fr-fr/sensitive-mature-themes/.

**4** The FYP is the first thing the user sees when he opens the TikTok app. It is a personalised feed of videos based on interests and engagement, which is algorithmically curated. https://support.tiktok.com/en/getting-started/for-you#.

ing on queer epistemology, these practices can also be studied as "algorithmic passing" strategies.

Passing refers to a person's ability to be considered a member of a hegemonic social group other than their own. Based on Sedgwick's work, Emmanuel Beaubatie (2019) considers passing as a strategy of adjustment for trans people in certain straight and cisgender spaces. This strategy, he says, is composite, unequal and shaped by subjective trajectories determined by sex, class and race. On social media, LGBT internet users wield a whole set of TikTok-compatible digital writings to stage themselves on the platform and become visible in a space that invisibilises them. To put it differently, writing oneself as gay online in a way requires passing the test of algorithmic recognition. Having these thoughts in mind, one can easily understand how heuristic it is to focus on a marginalised group to study how content circulates online and how content moderation works. It is because LGBT individuals are used to renegotiate their right to (in)visibility that content moderation policies and designs are heuristic for the study of sexual and gendered identities, and vice versa. Sexual and gendered identities challenge the way content moderation must be thought. Consequently, for this chapter – shedding light on the various tactics employed by LGBT tiktokers to bypass algorithms on social media – I intend to discuss to what extent looking at platform affordances can become a way to study digital infrastructures and online (in)virality.

# 2 Working on the Invisibilisation of Content: Suggested Method and Corpus

## 2.1 Access to Deleted Data and Black Box: Methodological Obstacles

While a great deal of papers in the social sciences have focused on finding the best way to address the virality of content published online, few have studied invisibilisation due to methodological issues. On the latter, many reports concerning content moderation and the use of AI systems have underlined the algorithmic black box as a hindrance to academic research (see Zuiderveen Borgesius 2018; Fergus et al. 2020; Défenseur des droits, CNIL 2020). Without access to data, how can we build up a corpus? How can we find evidence of content that no longer exists on the platforms under study? Finally, without knowledge of the computer code or criteria, how can we understand algorithmic moderation designs? This part addresses the methodological issues of collecting deleted data from social media and how one can build up a corpus when working on algorithms' impact on users.

With the opening of Twitter's API to researchers until May 2023[5], many studies on content moderation have been conducted based on quantitative-qualitative methods, using machine learning or automated language processing. Although the terms of use of the Twitter API do not allow us to collect content deleted from the platform, or to have access to the conditions of their deletion, I have developed, with Virginie Julliard and the CERES team[6], a computerised collection method to study the phenomena of online content invisibilisation. This survey, carried out in 2022, was based on an analysis of several tens of thousands of tweets collected over four months, and enabled us to put forward hypotheses on the causes of abusive censorship or the algorithmic and human modalities that were engaged in Twitter's moderation processes (see Grison et al. 2023). This method was made possible by the opening of the API and the possibility of carrying out keyword searches, *a priori* independently from the platform's recommendation outputs. This research also provided an opportunity to investigate the effectiveness of keyword moderation and, by extension, concluded on the relative effectiveness of collecting a corpus of content using keywords, as a means of studying content moderation.

On TikTok, as on other SNSs, massive data collection bypassing algorithmic recommendation was not possible until summer 2023[7]. Although scrapping and crawling methods did exist (see Zelle 2023) to come as close as possible to searches focusing on Twitter, they seemed even more complex to adapt when working on the designs themselves. My encounter with Taina Bucher's work on algorithmic imaginaries and her investigative methods reinforced my belief that technological opacity was no obstacle to understanding algorithms, whatever the platform under study. In the spirit of reverse-engineering, Bucher asks the following question when working on algorithms: "When confronted with the seemingly obscure and hidden, what are our methodological options?" (2016, 82). For her, it's less a question of gaining access to what might be inside the "black box" (see Pasquale 2015) but rather to consider them as machines that do, that make people do or even make people talk and therefore return to what they do. Fred Pailler (2019), in his

**5** When writing this chapter in summer 2023, Elon Musk had already announced the closing of Twitter's (now called "X") API to the public. European researchers are now waiting for the DSA (Digital Services Act) to become effective. According to article 40, researchers working on platforms' "systemic risks" will have the capacity to make an access request to big platforms' data. On this topic, see Julien Rossi's articles online: https://www.openedition.org/35318?lang=en.

**6** https://ceres.sorbonne-universite.fr.

**7** In anticipation of the DSA, ByteDance (the company owning TikTok) opened the API to researchers to access public and anonymous data. The corpus of data I am dealing with and the methodology I propose in this chapter are antecedent to this announcement.

PhD thesis on sexual cultures and computing devices, carried out an ethnographic survey of the web infrastructures as well. Interestingly, in the two studies cited above, to gather insights on the technical functioning of infrastructures and the way they impact users, both authors adopt particular, subjective and subjectivising positions and postures. In Bucher's case, this takes the form of interviews or collections of online discourse from internet users who talk about their (failed, strange, unexpected) interactions with algorithms, which Sophie Bishop (2019) calls "algorithmic gossip". In his work, Pailler explains how he created and animated various fake profiles for his investigation and how his approach was also nourished by multiple interactional resources with sex workers, by their sexual knowledge and savoir-faire, and in particular by an approach based on affects.

All in all, I see three key points that are particularly heuristic for the study of algorithmic designs and processes of online invisibilisation: First, the study of the technical functioning of a digital infrastructure does not necessarily require massive data collection and privileged access to the platforms; secondly, the study of how these devices work can be carried out from subjective postures (what I will call later, "savoirs d'expérience"); lastly, the study of how these platforms work also requires taking into account a few hijacked, unplanned or even botched uses. This is what I will later refer to as "platform affordances" (see Tiidenberg and van der Nagel 2020). Indeed, on this last point, it's worth looking at how both Bucher and Pailler study digital devices at the "frontier" between uses, knowledge, discourses, affects, representations and software. Specifically, Bucher builds knowledge about algorithms using the discourses and strategies of internet users to thwart the algorithmic recommendation they find "weird" (Bucher 2017). Thus, algorithmic opacity and platform non-planned uses are less obstacles than opportunities to think about hijacked postures of investigation.

## 2.2 A Situated Method Built on Experience-Based Knowledge to Study Invisibilisation

While confronted with the opacity of content moderation, I asked myself "what are [my] methodological options?" (by reference to Bucher 2016). Having no access to the content once it had been deleted, nor to the rules by which the algorithms operated, I did however have access to people's testimonies explaining that their content had been moderated. In other words, I had access to an entire pool of "savoirs d'expérience" about censorship online. It is this experience-based knowledge that forms the basis of my entire methodological protocol. It has been gathered over the past three years through exploratory interviews, press articles, a digital ethnography and collection of online testimonials, discussions with those close to me who

are affected. In short, my position has always been to start from the experience of censorship and the practices that stem from it and work backwards to develop hypotheses on the technical functioning of digital infrastructures.

In particular, as part of this investigation on TikTok, I drew inspiration from the walkthrough method as proposed by Light, Burgess and Duguay (2018). This approach provides a means for investigating TikTok's seemingly opaque algorithmic curation and understanding how the app interfaces with identity performances (see Duguay 2023). This is a situated approach of immersion in the application. It requires taking time, adopting a reflexive posture as one consults and collecting the data of interest. It also allows me to question the uses and representations that underpin the platform I'm studying. What's more, drawing on queer methodology, it's important to emphasise how these methodological protocols for digital inquiry and ethnography are also motivated by one's personal experience. In my case, the investment of my homosexual relatives is the mean by which I was able to access certain content (as they sent me content encountered on their respective FYPs before it was subject to algorithmic moderation).

## 2.3 Growing a Niche Algorithm to Study Content Recommendation and Invisibilisation

Since 2021, I have been working on "training" my TikTok recommendation algorithm: By spending time watching certain videos, conducting several keyword searches in the built-in search engine or by "liking" and "following", I've let TikTok's recommendation algorithm select the content that might be of interest to me regarding my topic of work. In the hours I spent using the app, I was able to collect several hundred videos by screen recording before storing them in albums in my phone gallery. At the same time, my gay friends and colleagues also sent me any gay sexual content they came across on their FYP. These dispatches were often the subject of collective discussions, astonishment and collaborative preanalyses, which I will attempt to document later. This method was therefore tested and fine-tuned over a long period of time (two years of digital fieldwork). The situated approach to corpus collection I developed is somewhat at odds with conventional approaches that study online virality using computational and informatics tools. Here, I've found it more effective to move away from the traditional diffusionist approaches to the study of virality and focus instead on the practices, attempts and strategies of "going viral" with invisibilised content, from situated and even marginalised trajectories and profiles.

However, it seems important to recall that this corpus is the last in a trilogy of corpora housed on Twitter and TikTok and collected with the aim of studying the impact of moderation devices on the digital writings of LGBT people and the way they incorporate algorithmic devices. The first was collected via computational collection methods on Twitter in the course of 2022. The second was collected using keywords via the search engines built into the TikTok and Twitter platforms. And the last corpus is the one I'm focusing on today, on the issue of sexuality.

The selected corpus is made of thousands of videos posted in French, English and Spanish by various tiktokers on the app. Insofar as many of the videos are similar, and with the need to delimit a stable corpus of videos for the purposes of techno-semiotic analysis, I stored 87 screen-recorded videos, each representing a single algorithmic bypassing practice. What these videos have in common is that they are of a sexual nature, either through the expression of some form of nudity, or because they contain sex education content, or because they are pornographic or intended to arouse sexual desire. I will now present the results of my analyses through the lens of platform affordances.

## 3 From Platform Affordances to Algorithmic Bypass Practices

The corpus I detailed above enabled me to identify the algorithmic circumvention practices that gay content creators put in place to disseminate sexual content on TikTok and, by extension, to formulate hypotheses on algorithmic intervention in moderation. I will focus more precisely on three key aspects: circumvention practices through text, image and sound.

The bypassing practices I'm going to present are to be considered through the prism of platform affordance theory. In computer language, affordances designate the uses preconceived by the designers of a software. In other words, what the technology offers and how it is used by internet users, for example. But this definition is both very broad and lacks precision in that it does not consider the way in which these affordances are co-constructed by the users and designers of these platforms. For example, the birth of the hashtag on Twitter resulted from the use of the hash word by internet users. This practice was then incorporated into the design of the application and became the referencing tag we know today. Tiidenberg and van der Nagel in *Sex & Social Media* (2020), and more broadly, other works in porn studies (i.e., Cao 2021), study affordances as negotiations between what is offered by platforms – the features – and the actual (mis)uses (hijacked or

"at the margins" of preconfigurations by design). In this context, platform affordances make it possible to study the gap between the technical possibilities offered and permitted to users by companies, and the actual ways in which they are used. In other words, affordances allow researchers to identify the gap between what platforms are designed for and how they are used for/as. In short, taking affordances into account represents a heuristic approach for studying (opaque) digital infrastructures.

## 3.1 Text Affordances

Within my corpus, several videos feature text that has been modified and/or altered by the content creators prior to publication. This alteration is carried out in different ways. One way is not to publish the text in the video's description bar but rather in the video's image and accompanied by emojis that mask certain characters of the hidden word. The assumption made by internet users is that algorithmic spotting is less obvious or systematic in the video image than in the description text. In one of the videos I collected, for example, the word "titties" is covered by a heart emoji. It can be assumed that this masking process makes it impossible to scan the image for algorithmic identification: only the human eye is able to identify the word chosen by the user. This means that content moderation algorithms would be unable to distinguish whether the focus of the video was the breasts of a cisgender woman. This process is reminiscent of the blurring and masking effects of the tabloid press, known for editorialising photos by concealing the identity of the persons photographed or areas of sex and violence. Thus, the content creator plays with the technical possibilities offered by the creative studio integrated into the TikTok software to transform her text into an image, and then this text-image into a mask.

Another way of hiding a keyword has become widely popular on TikTok since 2021. It's called algospeak[8]. This tactic consists in inventing neologisms to use keywords recognised as "hateful" or "illicit" by the algorithms, but in a roundabout way. Here are just a few examples from our corpus: The word gay becomes "g4y"; the term "faggot" becomes "f🥖"[9]; "dyke" becomes "dyk3"; "porn" is spelled "p0rn"; and "sex" becomes sometimes "seggs", sometimes "S3x", and so on. The possibilities are so infinite that it would be futile for the moderators to attempt to list them all – especially as they differ from one language to another. However

**8** https://www.washingtonpost.com/technology/2022/04/08/algospeak-tiktok-le-dollar-bean/.

**9** "f🥖" understood as "f-baguette" (to mimic the sound of the word "fag-got").

diverse they may be, they have the particularity of using the codes of "text message language", playing semiotically between spoken and written language and alternating the use of different keyboards – letters, numbers and emojis being used as if they belonged to the same linguistic code. This crypto language is not exclusive to TikTok and to sexual and gendered communities. In her work on activist crypto language on SNS, Alexandra Saemmer (2019) identified, for example, how these play with words operated as strategies of "in-communication with the machine" (2019, 129) in order to avoid the digital tracking of *Gilets Jaunes* on Facebook. In her paper on algorithmic control tactics, Emily van der Nagel (2018), inspired by the *Harry Potter* saga, refers to the strategy of "voldemorting" which she defines as follows: "The spell functions much like an online keyword search: web pages containing particular words or phrases are returned when someone enters them into a search engine. Voldemorting has become a key tactic of making things invisible while discussing them online" (van der Nagel 2018, 87). To extend van der Nagel's point, I think it's important to emphasise how corpus building through keyword entries is often not enough when working with content that is likely to be moderated. On the contrary, a situated collection will make it possible to identify content that anticipates censorship even before it is published, whereas a keyword-based collection would not have given me access to the corpus I am focusing on here. Let's also acknowledge the fact that this strategy of renaming things through neologisms is broadly used by hateful communities online, for toxic chat purposes and other far-right communities (see Kim, Wohn and Cha 2022; Weimann and Masri 2020).

## 3.2 Image Affordances

In the same article from 2018, van der Nagel discusses another tactic that users put in place to hijack the algorithmic functioning. Basically, she presents the way in which screenshots are a backdoor means of publishing content without promoting them. Indeed, some users publish screenshots of posts rather than retweeting or quoting them, to avoid generating traffic or engagement. In my corpus, screenshots are often a clever way of republishing content that has already been moderated. Often using TikTok's "green screen" feature, creators place themselves in front of the camera and before the screenshot of the deleted content (which forms the green screen) and thus put back into circulation content that had previously been deleted from the platform. In this way, they circumvent algorithmic censorship by republishing content in a new format.

There are a multitude of tactics for influencing algorithmic referencing or slipping through the cracks of moderation. Within my corpus of sexual videos, I

was able to classify them in nine different types of strategies. Nonetheless, this typology remains non-exhaustive as new ones are appearing every day, due to the evolution of cultures, digital languages and content moderation rules and designs' adjustments. The following table[10] summarises the typology of tactics imagined by LGBT users to expose sexual content on TikTok (see Tab. 1).

**Tab. 1:** Tactics imagined by LGBT users to expose sexual content on TikTok.

| Tactic | Explanation |
|---|---|
| Screenshot | Usually, the repost of a content that has already been deleted from the platform. It is usually framed in the bigger picture. |
| Green screen | Sometimes associated with the screenshot: the content creator appears in front of the screen with a screenshot behind him. The person in front of the camera also partially hides the image behind. |
| Hinting | The sexual act is suggested off-camera or simply gestured. |
| Blurring | The sexually explicit area is hidden by blurring or by adding a mask (which may take the form of an emoji or a banner). |
| Flashing | From slang "to flash someone". It consists in showing very quickly one's naked (part of the) body and then hide it again. On TikTok, it usually appears as a nude that is shown for only half a second in the middle of a slower video or slideshow. The appearance of the photo is so fleeting that the sexual nature of the photo does not seem to be identified by the machine or the moderator. |
| Shadowing | The sexual act is shown through shadows or body shapes. |
| Isotopia | The sexual nature of the TikTok video is signified by the reference to the accumulation of sexual symbols and fetishes (socks, hoods, leather harness, milk etc.) |
| Metonymy | The sexual nature of the video is suggested or shown by a sexual object or symbol other than a naked body part. For instance, I found in the corpus a lot of photographs of semen deposited on a surface) |
| "Use this filter" | The naked part of the image is transformed by AI to appear in a cartoon or another shape, using TikTok filters. |

These circumvention strategies lead to several hypotheses about how sexuality is moderated online, and whether or not moderation algorithms prove to be effective. First, these practices seem to confirm that sex is moderated by the presence of nudity in the image. Moreover, this nudity seems to be identified by the algo-

**10** Due to copyright and permission to reproduce images, the screenshots have been removed from the table.

rithms through pixels of skin or the recognition of sexual attributes (penis, breast, vagina, etc.) (see Gehl et al. 2017). A video filming semen is therefore not necessarily moderated when the phallus is absent from the frame. Indeed, in the example labelled "metonymy", the photo of sperm appears after an image of a milk brick and where the viewer might understand the shift towards sexual content, the machine only considers sperm as a simple liquid and does not acknowledge the sexual nature of the video. Finally, beyond the visual object that triggers moderation, the practice of "flashing" is extremely valuable to consider. In this video, the sexual image is quite explicit in that it shows the bodies of internet users or porn stars naked, during a sexual act or in suggestive positions. If the content creators disseminating this image had wanted to publish the video without the flashing tactic, the image recognition algorithm would have deleted it before it was even disseminated. So, the only explanation I can see to justify its online circulation is the display time of the pornographic image: Indeed, the image in question is only visible in the blink of an eye. Does this mean that the pornographic image recognition algorithms employed by TikTok require a certain amount of time to identify the pixels of nudity? If so, here's an interesting example of how thinking from users' "savoirs d'expérience" and tactics can shed light on how opaque *dispositifs* technically work.

Funnily enough, it is also worth highlighting how these algorithmic bypassing tactics can become a guarantee for making content go viral. Take flashing videos, for example: I wondered why flashing videos were so popular in my corpus (indeed they cumulated lots of views and likes for each video collected). In fact, it's because this content need to be watched several times to spot the nudity image that appears very briefly in the video. Also, internet users who want to "screen" the image in question to watch it for longer are often forced to watch the video several times before spotting the right moment. These repeated viewings, the unexpected nature of the video, the screenshots or even the sharing of this content with one's relatives promote engagement with the video and therefore boost their ranking and recommendation among other internet users. As it has become a trend like any other, these flashing videos have ended up becoming popular content, even though they are contrary to the community guidelines of the platform and even though content moderation systems are supposed to prevent their massive circulation. Note that several studies have also pointed out the hypocrisy of the recommendation systems of other SNS such as Instagram, where swimsuit photos are heavily pushed by the algorithm even though the company advocates a puritanical vision of sexuality. This is particularly the case for the exposure of women's bodies, which are, for instance, both censored and sexualised on Instagram (see Lesage 2022).

## 3.3 Sound Affordances

At first, TikTok was supposed to be an extension of Musical.ly, a singing and dancing platform aimed at teenagers. Now, among other things, TikTok has become a digital space for musical artists to create new content and collaborate directly other creators using unique platform features (see Kaye et al. 2021; 2022). In this context, TikTok has also been considered a transmedia platform that enables new digital activism approaches using music. This means that users have initiated a phenomenon that consists in the appropriation of an audio meme to express their emotions, their beliefs and fights (see Vizcaíno-Verdú and Aguaded 2022). In my field, sound is indeed used to denounce the censorship that certain communities are victims of and to inveigh against the platform (see Grison and Julliard 2024)[11].

This dissemination is obviously encouraged by what the platform's features allow and encourage: TikTok indeed offers the possibility to extract the sound from a video to be able to recreate content with the same soundtrack, making its diffusion easier. The use of certain "trending" sounds is often a guarantee of visibility for one's video. It should also be noted that beyond this feature, which encourages the creation of content, sound acts as a hashtag on TikTok. It is therefore a referential that allows all content that use it to be classified and listed in the same place. Indeed, on TikTok, users can consult all the content posted using a specific sound when clicking on the link below a video. This feature is therefore used as a library of content using the same sound by users.

For this chapter, I have especially worked on how music and sound are also used on TikTok to give a sexual character to a video or to reference content with a sexual or erotic character. While the dissemination of video excerpts from pornography sites is prohibited on TikTok, the use of a sexual sound is not – or, at least it does not seem easily detectable by algorithms. Thus, a certain number of videos whose image does not *a priori* present any mention of sexuality are sexualised by the addition of a soundtrack. For example, one of my gay relatives sent me a video of Tom Holland, an English actor, taken from one of his interviews with the celebrity press. The sound in question, since removed by the platform, is an excerpt from a pornographic scene in which individuals can be heard moaning in pleasure and having sex. The sound thus participates in a semiotisation of the sexual act by giving a sexual character to the video and to the actor, whose image was initially devoid of it.

**11** For example, https://www.tiktok.com/music/Dear-TikTok-Spoken-Word-6949993320974207750?lang=fr&is_copy_url=1&is_from_webapp=v1.

However, another use of sound can be in the referencing of sexually explicit content. When the same sound is used by several internet users to disseminate sexual videos, they are in fact listed in the same place. These sounds then become niche libraries of sexual content on a social network. Internet users therefore play with the platforms' features to create usages at the margins of what is authorised or thought of by the developers. These usages then transform into community practices that influence the way these applications are used.

It is interesting to note how these languages (such as algospeak) or these tactics are constitutive of new online writing practices (see Souchier et al. 2019), which are themselves constitutive of semiotic and political communities (see Julliard and Saemmer 2022)[12]. In fact, in three years of work in this field, I have observed how neologisms, sound semiotisations, image alterations, etc. have gradually become vernacularised within online sexual and gendered communities. In a sense, this echoes hypotheses that in SNSs, communities generally do not pre-exist their writing practices and publications (see Julliard 2022). Here, a semiotic community is structured around the contestation of mechanisms of invisibilisation of LGBT content or illicit content publishing practices on the platforms. These practices, considered as platform affordances, capable of juggling with the limits of usage preconceived by designers, allow for an exploration of the possibilities of writing oneself, expressing oneself and being seen in and through the digital writing devices that are SNSs.

## Conclusion

This case study focused on the shadowbanning of LGBT content on TikTok and aimed to examine how marginalised communities navigate and bypass censorship on online platforms. The study proposed a new approach to investigating online dereferencing or censorship by not only collecting the content itself, but also analysing the practices used to circumvent censorship. Ultimately, the research revealed the tactics used by individuals and sexual and gendered communities to overcome algorithmic censorship and shed light on the way LGBT users negotiate their visibility online.

I would like to conclude on the idea of virality as negotiation, which runs throughout this chapter. I propose to discuss how content can "go viral" or how content are "being censored" through the lens of negotiation at two different lev-

**12** For example, algospeak has become a means of identifying oneself as belonging to "Gay TikTok" rather than a means of avoiding censorship today, insofar as the use of keywords like "gay" is no longer punished as much as it was before 2022 on TikTok.

els and in regard with content moderation *dispositifs*. First, I believe that virality consists in a negotiation between the intentions and digital writing skills of internet users, as well as with the technical features offered and constrained by the platforms' designs. Additionally, I believe that virality can be studied as a cultural negotiation between the normative editorial frameworks of generalist websites (such as the SNSs like TikTok) and the cultural specificities of subaltern or hegemonic communities. These negotiations occur repeatedly and often in a conflictual manner, influenced by the rules set by platforms, the technologies used and the communities involved.

Virality of content as negotiation naturally has an impact on the way we write about ourselves as queers, dykes, trans, Blacks, Arabs, feminists and so on. Thus, it plays as much a part in their sudden (hyper)visibility as in their segregation or in their cultural transformation throughout time (through the birth of digital crypto languages, for example). In many ways, these negotiations are central to understanding how content is made (in)visible online and to what extent some communities are more impacted than others by content moderation designs.

## References


Are, Carolina. "How Instagram's Algorithm Is Censoring Women and Vulnerable Users but Helping Online Abusers." *Feminist Media Studies* 20, no. 5 (2020): 741–744. https://doi.org/10.1080/14680777.2020.1783805.

Armangau, Yael, and JulienFigeac. "Les formes de la solidarité sociale dans les groupes facebook trans." *Réseaux* 6, no. 242 (2023). https://www.cairn.info/revue-reseaux-2023-0-page-I2.htm.

Badouard, Romain. *Les nouvelles lois du web. Modération et censure*. Paris: Seuil, 2020.

Beaubatie, Emmanuel. "L'aménagement du placard: Rapports sociaux et invisibilité chez les hommes et les femmes trans' en France." *Genèses* 114, no. 1 (2019): 32–52. https://doi.org/10.3917/gen.114.0032.

Benjamin, Ruha. *Race after technology: abolitionist tools for the new Jim code*. Medford, MA: Polity, 2019.

Berlant, Lauren, and Michael Warner. "Sexe en public." *Questions de communication* 33 (2018): 111–133. https://doi.org/10.4000/questionsdecommunication.12204.

Bishop, Sophie. "Managing Visibility on YouTube through Algorithmic Gossip." *New Media & Society* 21, no. 11-12 (2019): 2589–2606. https://doi.org/10.1177/1461444819854731.

Bowker, Geoffrey C., and Susan Leigh Star. *Sorting Things out: Classification and Its Consequences*. Cambridge, MA: MIT Press, 2008.

Bucher, Taina. "Neither Black Nor Box: Ways of Knowing Algorithms." In *Innovative Methods in Media and Communication Research*, edited by Sebastian Kubitschko and Anne Kaun, 81–98. Cham: Springer International Publishing, 2016. https://doi.org/10.1007/978-3-319-40700-5_5.

Bucher, Taina. "The Algorithmic Imaginary: Exploring the Ordinary Affects of Facebook Algorithms." *Information, Communication & Society* 20, no. 1 (2017): 30–44. https://doi.org/10.1080/1369118X.2016.1154086.

Cao, Ruepert Jiel Dionisio. “Amateur Porn in Filipino Twitter Alter Community: Affordances, Commodification, Ghettoization, and Gay Masculinity.” *Media International Australia* 179, no. 1 (2021): 52–65. https://doi.org/10.1177/1329878X211002845.

Chan, Ngai Keung, Chao Su, Chris, and Alexis Shore. “Shifting platform values in community guidelines: Examining the evolution of TikTok’s governance frameworks.” *New Media & Society* 0, no. 0 (2023). https://doi.org/10.1177/14614448231189476.

Cho, Alexander. “Default Publicness: Queer Youth of Color, Social Media, and Being Outed by the Machine.” *New Media & Society* 20, no. 9 (2018): 3183–3200. https://doi.org/10.1177/1461444817744784.

Damian-Gaillard, Béatrice. “L’économie politique du désir dans la presse pornographique hétérosexuelle masculine française.” *Questions de communication*, no. 26 (2014): 39–54. https://doi.org/10.4000/questionsdecommunication.9224.

Défenseur des droits. “Algorithmes : prévenir l’automatisation des discriminations.” CNIL, 2020.

Dell’Omodarme, Marco. “Pour une épistémologie des savoirs situés: De l’épistémologie génétique de Jean Piaget aux savoirs critiques.” PhD thesis. University Paris 1 Panthéon-Sorbonne, 2014.

Despontin Lefevre, Irène. “Stratégies de communication et pratiques militantes dans le mouvement féministe en France au début du 21ème siècle. Étude de cas du collectif #NousToutes (2018–2021).” PhD thesis. University Paris-Pantheon Assas, 2022.

Duguay, Stefanie. “TikTok’s Queer Potential: Identity, Methods, Movements.” *Social Media + Society* 9, no. 1 (2023): 205630512311575. https://doi.org/10.1177/20563051231157594.

Fergus, Ryan, Fritz, Audrey, and Daria Impiombato. “Tiktok and WeChat : Curating and controlling global information flows.” ASPI, International Cyber Policy Centre, 2020.

Gagnon, John H. “Les usages explicites et implicites de la perspective des scripts dans les recherches sur la sexualité.” *Actes de la recherche en sciences sociales* 128, (1999) : 73–79. 10.3406/arss.1999.3515.

Gehl, Robert W., Moyer-Horner, Lucas, and Sara K. Yeo. “Lucas Training Computers to See Internet Pornography: Gender and Sexual Discrimination in Computer Vision Science.” *Television & New Media* 18, no. 6 (2017): 529–547. https://doi.org/10.1177/1527476416680453.

Gillespie, Tarleton. *Custodians of the internet: platforms, content moderation, and the hidden decisions that shape social media*. New Haven: Yale University Press, 2018.

Gillespie, Tarleton, Aufderheide, Patricia, Carmi, Elinor, Gerrard, Ysabel, Gorwa, Robert, Matamoros-Fernández, Ariadna, Roberts, Sarah T., Sinnreich, Amram, and Sarah Myers West. “Expanding the debate about content moderation: scholarly research agendas for the coming policy debates.” *Internet Policy Review* 9 no. 4 (2020). https://doi.org/10.14763/2020.4.1512.

Grison, Thibault, and Virginie Julliard. “Les enjeux de la modération automatisée sur les réseaux sociaux numériques : les mobilisations LGBT contre la loi Avia.” *Communication, technologies et développement* 10 (2021). https://doi.org/10.4000/ctd.6049.

Grison, Thibault, Julliard, Virginie, Alié, Felix, and Victor Écrement. “La modération abusive sur Twitter: Étude de cas sur l’invisibilisation des contenus LGBT et TDS en ligne.” *Réseaux* 237, no. 1 (2023): 119–149. https://doi.org/10.3917/res.237.0119.

Grison, Thibault, and Virginie Julliard. “L’IA entre censure et terreau de nouvelles formes sémiotiques dans la modération des contenus LGBT sur Twitter et TikTok”. In *Intelligence artificielle, culture et médias*, edited by Guèvremont, Véronique, and Colette Brin, forthcoming. Laval: Presses de l’université de Laval, 2024.

Jahjah, Marc. “‘T’es intelligent pour un arabe !’ Auto-ethnographie d’un corps colonisé: Une épistémologie du mezzé libanais.” *Itinéraires* 2021, no 3 (2022). https://doi.org/10.4000/itineraires.11748.

Jean, Aurélie. *De l'autre côté de la machine: voyage d'une scientifique au pays des algorithmes*. Paris: Éditions de l'Observatoire, 2019.

Jouët, Josiane. *Numérique, féminisme et société*. Paris: Presses des Mines, 2022.

Julliard, Virginie. "Communauté politique, sémiotique, émotionnelle: Ce que la circulation des images révèle de la structuration de la mobilisation anti-genre sur Twitter." *Communication & langages* 212, no. 2 (2022): 131–153. https://doi.org/10.3917/comla1.212.0131.

Julliard, Virginie, and Alexandra Saemmer. "Un regard situé pour étudier les communautés interprétatives et émotionnelles." *Communication & langages* 212, no. 2 (2022): 21–32. https://doi.org/10.3917/comla1.212.0021.

Kaye, Bondy Valdovinos. "Please duet this: collaborative music making in lockdown on TikTok." Networking Knowledge: *Journal of the MeCCSA Postgraduate Network* 15, no. 1 (2022). https://ojs.meccsa.org.uk/index.php/netknow/article/view/654.

Kaye, Bondy Valdovinos, Rodriguez, Aleesha, and Patrick Wikstrom. "You Made This? I Made This: Practices of Authorship." *International Journal of Communication* 15 (2021), 3195–3215.

Kim, Jaeheon, Donghee, Yvette Wohn, and Meeyoung Cha. "Understanding and identifying the use of emotes in toxic chat on Twitch." *Online Social Networks and Media* 27 (2022). https://doi.org/10.1016/j.osnem.2021.100180.

Lesage, Agate. "Instagram et la censure des corps sexisés." *Mouvements* 112, no. 4 (2022): 147–156.

Light, Ben, Burgess, Jean, and Stefanie Duguay. "The Walkthrough Method: An Approach to the Study of Apps." *New Media & Society* 20, no. 3 (2018): 881–900. https://doi.org/10.1177/1461444816675438.

Monea, Alexander. *The digital closet: how the internet became straight*. Cambridge, MA: MIT Press, 2022.

Nagel, Emily van der. "'Networks That Work Too Well': Intervening in Algorithmic Connections." *Media International Australia* 168, no. 1 (2018): 81–92. https://doi.org/10.1177/1329878X18783002.

Noble, Safiya Umoja. *Algorithms of oppression: how search engines reinforce racism*. New York: New York University Press, 2018.

Paasonen, Susanna, Jarrett, Kylie, Light, Ben, and Florian Vörös. "Puritanisme sexuel et capitalisme numérique." *Revue Française de Socio-Économie* 25, no. 2 (2020): 167–174. https://doi.org/10.3917/rfse.025.0167.

Pailler, fred. "Les affects classifiés : numérique et médiations sexuelles." PhD thesis. Université de Nantes, 2019. https://hal.archives-ouvertes.fr/tel-03236631/document.

Pasquale, Frank. *The Black Box Society, the Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press, 2015.

Saemmer, Alexandra. "Le parler fransais des Gilles et John. Enquête sur les crypto-langages militants au sein des plateformes." *Hermès* 84, no. 2 (2019): 127–133. https://doi.org/10.3917/herm.084.0127.

Sedgwick, Eve Kosofsky. *Epistemology of the closet*. Berkeley: University of California Press, 1990.

Seering, Joseph. "Reconsidering Self-Moderation: The Role of Research in Supporting Community-Based Models for Online Content Moderation." *Proceedings of the ACM on Human-Computer Interaction* 4, (2020): 1–28.

Souchier, Emmanuël, Candel, Etienne, Gomez-Mejia, Gustavo, and Valérie Jeanne-Perrier. *Le numérique comme écriture: théories et méthodes d'analyse*. Malakoff: Armand Colin, 2019.

Tiidenberg, Katrin, and Emily van der Nagel. *Sex and Social Media*. Bingley: Emerald publishing, 2020.

van der Nagel, Emily. "Competing platform imaginaries of NSFW content creation on OnlyFans." *Porn Studies* 8, no. 4, (2021): 394–410. 10.1080/23268743.2021.1974927.

Vizcaíno-Verdú, Arantxa, and Ignacio Aguaded. "#ThisIsMeChallenge and Music for Empowerment of Marginalized Groups on TikTok." *Media and Communication* 10, no. 1 (2022). https://doi.org/10.17645/mac.v10i1.4715.

Vörös, Florian. *Désirer comme un homme: enquête sur les fantasmes et les masculinités*. Paris: La Découverte, 2020.

Weimann, Gabriel, and Natalie Masri. "Research Note: Spreading Hate on TikTok". *Studies in Conflict & Terrorism* 46 (2020): 1–14. DOI: 10.1080/1057610X.2020.1780027.

Zelle, Lars. "Let's Talk about TikTok – A Web Scraping Tool for Social Science Research." Bachelor thesis. Heinrich-Heine-University Düsseldorf, 2023. DOI:10.13140/RG.2.2.10710.11843.

Zuiderveen Borgesius, Frederik. "Discrimination, intelligence artificielle et décisions algorithmiques." Strasbourg: Conseil de l'Europe, 2018. https://rm.coe.int/etude-sur-discrimination-intelligence-artificielle-et-decisions-algori/1680925d84.