

NETWORKED INTIMACIES: SOCIAL MEDIA POLICIES AND REGULATION ON ADULT  
CONTENT

WILLA MICHAELSEN

This dissertation was submitted in part fulfilment of requirements for the degree of MSc  
Information and Library Studies

DEPARTMENT OF COMPUTER AND INFORMATION SCIENCES  
UNIVERSITY OF STRATHCLYDE  
GLASGOW, SCOTLAND

AUGUST 2019

## DECLARATION

This dissertation is submitted in part fulfilment of the requirements for the degree of MSc of the University of Strathclyde.

I declare that this dissertation embodies the results of my own work and that it has been composed by myself. Following normal academic conventions, I have made due acknowledgement to the work of others.

I declare that I have sought, and received, ethics approval via the Departmental Ethics Committee as appropriate to my research.

I give permission to the University of Strathclyde, Department of Computer and Information Sciences, to provide copies of the dissertation, at cost, to those who may in the future request a copy of the dissertation for private study or research.

I give permission to the University of Strathclyde, Department of Computer and Information Sciences, to place a copy of the dissertation in a publicly available archive.

(Please tick) Yes [  ] No [  ]

I declare that the word count for this dissertation (excluding title page, declaration, abstract, acknowledgements, table of contents, list of illustrations, references and appendices is **19,417**).

I confirm that I wish this to be assessed as a Type  1  2  3  4  5 Dissertation (please circle).

Signature:

Date:

## ABSTRACT

This dissertation is an extended literature survey and critical discourse analysis of social media policies about adult content. It examines how and by whom adult content is regulated using Lessig's framework and the methods of critical discourse analysis. I focussed on what I consider to be the major social media platforms in the English-language world: Facebook, Twitter, Tumblr, Instagram, and YouTube. I chose these specifically because of their size and relevance to the daily life of the average person. Lessig introduced the four modalities of Internet regulation in his 2000 book, *Code is Law: norms, laws, architecture, and the market*. I will define each one to be relevant to my purposes. Then I will use the four to structure my analysis and categorise the social media policies by which modes they are exemplifying. Sometimes the modes come into conflict with each other in ways that will be analytically useful. Norms are the ways communities self-regulate, through established acceptable behaviours. These can be explicit or implicit, and can be inflicted externally and adopted by the community, or built from the ground up within the community. Norms can change over time, and conflict results from not knowing or disregarding the norms, or when individual communities interact and try to reconcile their divergent norms. Laws regulate the Internet by threatening consequences for undesirable behaviour, on the part of Internet users and its architects. Architecture is all the ways the Internet is constructed, and the affordances of those design and code decisions. Its regulation can often be unseen, but architecture shapes the very spaces we enter online, let alone the conversations we have once we are there. The market inflicts regulation because social media companies exist as profit-driven ventures within capitalism. Critical discourse analysis is the examination of the ways language reflects and reinforces social organisation. Critical discourse analysis will be used to examine if and how adult content policies (re)produce societal restrictions on women, LGBTQIA people, people of colour, and sex workers. By controlling the ways in which marginalised people can speak online, their marginalisation is maintained.

## ACKNOWLEDGEMENTS

This dissertation would not have been possible without the advocacy, activism, and authorship of sex workers and sex worker-led organisations. I encourage anyone reading this to support their efforts. SWARM ([swarmcollective.org](http://swarmcollective.org)) is an incredible grassroots group based in Glasgow. DecrimNow ([decrimnow.org.uk](http://decrimnow.org.uk)) is UK-wide, with sister decriminalisation efforts in the US like DecrimNY ([decrimny.org](http://decrimny.org)). You may have preconceptions about sex work and its dangers, but please consider first listening to the voices of actual sex workers regarding their needs. We don't have to agree, but harm reduction should be at the heart of our activism.

I would like to thank my dissertation supervisor, Dr David McMenemy, for his guidance and advice. He pointed me towards the work of Lessig as an analytical framework and the methodology of critical discourse analysis. To my parents, Susan and Joel, thank you for your endless support. Thanks also to my American contingent, Stew, Nick, and Will, and to my Scottish contingent, Bridget and Daniel.

# TABLE OF CONTENTS

1. Introduction .....	1
1.1 Background and Context.....	2
2. Methodology .....	4
2.1 Researcher Motivation .....	4
2.2 Research Questions.....	4
2.3 Process.....	5
2.4 Literature Review: Themes and Focus.....	7
2.5 Lessig’s Regulatory Modes.....	7
2.6 Critical Discourse Analysis.....	8
3. Literature Review: The Law .....	9
3.1 The Communications Decency Act.....	9
3.2 Child Online Protection Act.....	11
3.3 Stop Enabling Sex Traffickers Act .....	12
3.4 Sex Worker Response to SESTA-FOSTA and the Rescue Industry.....	12
4. Literature Review: Networked Publics.....	15
4.1 The Political Public.....	16
4.2 The Cultural Public.....	17
4.3 Activist Counterpublics.....	18
4.4 Lessig’s Regulation and the Public Sphere .....	19
5. Literature Review: Content Moderation.....	20
5.1 Flagging and Reporting.....	22
5.2 Human Moderation versus Automatic Content Detection .....	23
5.3 Free Speech Tensions.....	24
6. Literature Review: Sex and Sexuality Online .....	25
6.1 Porn: For Who, By Who? .....	25
6.2 Feminist Pornographies .....	27
6.3 Networked Queerness .....	28
6.4 Sex Work Online .....	30
6.5 Sex Workers Organizing Online .....	31
7. Social Media Policy Analysis .....	32
7.1 Laws .....	33
7.2 Norms.....	38
7.3 Architecture .....	40

7.4 Markets.....	43
8. Conclusion .....	44
8.1 How do social media companies regulate adult content? .....	45
8.2 Can we see each mode of Lessig's regulation working on and through social media policies?.....	45
8.3 What effect does regulating adult content have on marginalized people? .....	45
9. Reflections and Recommendations.....	46
References.....	47
Appendix: Social Media Policies.....	I
Facebook 2018.....	I
Facebook 2019.....	IV
Instagram 2019 .....	IX
Tumblr 2017 .....	XI
Tumblr 2018 .....	XV
Twitter 2019.....	XVIII
YouTube 2019.....	XXIII

# 1. Introduction

There is a research gap in examining the long-term relationship between platforms and adult content. People have been engaging in social networking online for two decades; it is time to take an historical view. Through synthesising research into social media as publics and counterpublics, Internet law, and the methods of critical discourse analysis, I will examine how social networks both silence and permit intimacies and social organising. Platforms are not neutral spaces, but are instead revenue-generating advertising delivery methods. Therefore, despite their stated views, social networks are not driven by ideological commitment to freedom of expression (Tufekci, 2017). There is a general perception that the Internet used to be a place for free exchange of ideas but now is ruled by corporations that silo groups of people into echo chambers. This dissertation seeks to apply rigour to this perception, to complicate the narrative.

The central research question is thus: how have social networks regulated adult content over time? By answering this question, we can perhaps draw larger conclusions about shifting cultural attitudes and think critically before accepting the mechanisms of regulation.

This dissertation will examine how and by whom adult content is regulated using Lessig's framework and the methods of critical discourse analysis. We can classify regulation into the four modes of norms, laws, architecture or code, and the market (Lessig 2000). Norms are the behavioural standards to which one must adhere, established and enforced both by the platform's creators and one's fellow users. Code regulates by monitoring, shaping, and restricting cyberspace and its users' actions. For instance, as you navigate the Internet, your actions are recorded by the cookies on your browser and your location tracked by your IP address. When you want to watch a video that is only intended to be displayed to people in certain countries and are unable to, this is an example of regulation through code. This example also illuminates regulation through the market. Ways of tracking identity online were developed to improve the functioning of commerce. Their existence today, as Lessig forecast in 2000, have dramatically changed the regulability of the Internet. However, he also warned of the dangers of letting the free market shape our online spaces: "when commercial interests determine the architecture, they create a kind of privatised law" (p. 77). To what extent can we see each mode of Lessig's regulation working on social networks?

## 1.1 Background and Context

Facebook and Twitter have set out to undertake the sort of project never before imaginable in human history. A global marketplace, the town square writ large, where anyone anywhere can access a constant stream of communication and ideas with only a phone and a data connection. Of course, they have also managed to generate billions of dollars by harvesting and selling the personal information of their users, to advance surveillance to levels of reach and constancy envisioned only in dystopian science fiction. This is our world today--both wildly utopian and grimly dystopian. And as is the case with much utopian fiction, there is a dark side to what seems like free and open global connectivity, the marketplace of ideas. Hosting a truly free and open platform is impossible. Any attempts to do so become immediately inundated with spam. These spaces without rules of content regulation also become the online safe havens for extremist groups to recruit, plan, and marinate in their hatred. Also, Facebook and Twitter are publicly traded companies who must establish rules of conduct to remain palatable to investors. Despite their rhetoric--Mark Zuckerberg's famous motto "move fast, break things"--companies are really rather conservative and not biased towards action.

So if Facebook's "goal is to create a safe and welcoming community for the more than 2 billion people who use Facebook around the world, across cultures and perspectives" (*Community standards*, 2018), and keep investors on board, and dodge legal action or civil liability, then regulate they must. This dissertation will explore regulation in the more specific realm of adult content. When measuring what sort of content is allowed, there is no objective standard. Cultural attitudes regarding nudity, obscenity, and sexuality vary globally; cultural attitudes are not monolithic even within countries. Through examining the historical trends of Internet adult content moderation, we can track the ways platforms have wrestled with these ideas over time, and to what end. Can we identify specific cultural tipping points that led to policy changes? What laws regulate this area of the internet--decency laws, sex work laws, laws to protect children--and how have platforms adapted to comply with them? What are some implications of content moderation on freedom of expression?

In November of 2018, Apple expunged the app of microblogging platform Tumblr from the App Store, due to reported instances of child sexual exploitation (CSE) (Porter, 2018). After an initial pornography purge failed to regain admission to the App Store, Tumblr suddenly announced they were banning all explicit content in December 2018, including all "female-presenting



nipples” and “any content...including illustrations, that depicts sex acts” (*Community guidelines* 2019). Considering approximately 20% of Tumblr’s traffic had been consumers and producers of explicit media (Perez, 2017), questions were raised both about the wisdom of this decision and where that audience would go. Another recent instance of these policies in the public view is Twitter’s decision to ban ‘nonconsensual intimate media’ ie, revenge porn and, even more recently, the wide new world of deep-fakes. It is important to take a critical lens to these policies, because neutrality can only exist when speaking from a position of power. Neutrality is a fiction that benefits the empowered. Any rule that is designed to be universally applied can be wielded as a cudgel to harm those already at a disadvantage. To some, deleting all explicit content may seem unimportant or irrelevant to their experience on the platform. However, queer romantic content is generally deemed more explicit than heterosexual content. We can also see that the lives of sex workers and adult performers are criminalised even when they are ‘off the clock’, as sex workers have recently been banned from most online payment services like PayPal, Venmo, etc. So to others, these policies are restricting their ability to participate in the online world. This is harmful because of social media’s power for organising activism and sharing crucial health, sex, and sexuality information. How does regulation restrict the abilities of marginalised groups to participate fully in social networks?

The definition of social media has changed over the years to reflect its evolution and the shifts in our relationship and understanding. One generally accepted definition of social media is as “a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content” (Kaplan and Haenlein, 2010, p. 60). I would add to the definition that they also facilitate the monetisation of user-generated content. Social media broke down some of the barriers inherent to traditional media--its one-sidedness, the privileged position of the ones afforded public speech. Yet, the structure of social networking sites also reflects the offline world. Critical discourse analysis shows us there is power in discourse (Fairclough, 2001) and that “discourse is ideological in so far as it contributes to sustaining particular relations of power and domination” (p. 126). The online world is not different from the ‘real’ world in the ways it produces and maintains inequalities. Social media “form[s] a new online layer through which people organise their lives...influenc[ing] human interaction on an individual and community level, as well as on a larger societal level, while the worlds of online and offline are increasingly interpenetrating” (Van Dijck, 2013, p. 4). Lawrence Lessig’s theoretical framework of Internet regulation--norms, laws,

architecture and the market--allows us to examine the forces that shape online discourse internally and externally, visibly and not (2000).

## 2. Methodology

### 2.1 Researcher Motivation

I was inspired to undertake a study on adult content and social media due to my personal experiences online. As a queer, Internet-savvy millennial, I have been aware of these issues through observation and conversation amongst some of the online communities in which I am a member. I wanted to apply a more analytical awareness to what had been a general sense of restrictions and loss over time. Lawrence Lessig codified the four modes of Internet regulation that serve as my analytical framework: norms, laws, architecture, and the market. Looking at how social media companies have tried to regulate adult content over time, perhaps we can build a greater understanding of this relationship between platform and user and how different communities have both arisen and dissolved on social media. After an initial boom of freedom of expression, and a wonderful sense of possibility many of us felt in the earlier days of the Internet, I have been watching over the last several years as Internet culture has appeared to become corporate-controlled, regressive, and conservative. Through a wide review of the literature, I want to examine why I feel this way.

### 2.2 Research Questions

1. How do social media companies regulate adult content?
2. How has this regulation changed over time?
3. What incidents, cultural shifts, and changes in law led to policy changes?
4. Can we see each mode of Lessig's regulation working on and through social media policies?
5. Where and to what extent?
6. What effect does regulating adult content have on marginalised groups?
7. Who is most affected, and how?

## 2.3 Process

This dissertation is an extended literature survey, examining the text of social media policies and surrounding literature. Kucan (2011) asserts the importance of the rhetorical literature review, in that the literature review should compellingly argue the *raison d'être* for the research the dissertation student is undertaking and “the impetus for a current study is well described and the rationale is well grounded” (p. 230). I chose the literature review over a survey or interviews because I am attempting to understand regulation from a top-down perspective, in that I want to look at how companies, lawmakers, and coders inflict regulation on the inherently unstructured space of the Internet. Some of the literature does dive into user experiences and interpretations of regulation, which adds an interesting element of examining how and why those expectations between regulator and regulated can be mismatched. Usually, these mismatches lead to friction that results in change, whether it is policy change or user behavioural change. By doing this as a literature review, I am combining inductive and deductive reasoning. Working from my own perspectives and accounts of specific incidents, I am looking to draw broader conclusions about social media at large. By taking the social media policies all together and using them to analyse the same incidents, I take a more deductive approach.

I collected a wide array of social media terms of service, community guidelines, and rules, regarding nudity, sexuality, and adult content using the Wayback Machine, current policies, and Internet reporters' work on leaked documents. I focussed on what I consider to be the major social media platforms in the English-language world: Facebook, Twitter, Tumblr, Instagram, and YouTube. I chose these specifically because of their size and relevance to the daily life of the average person. Each platform has multiple entry points for each possible device, including web pages and apps. Kaplan and Haenlein define social media by its ability to facilitate “the creation and exchange of user generated content” (2010, p. 60). For my purposes, a workable definition also includes the awareness that ‘the creation and exchange of user-generated content’ is a revenue-generating enterprise for social media companies. They have created a value-added product by combining elements of a multimedia entertainment service and individual-to-individual communication. This creates value and incentive for the users. Their content and data then in turn become value to the social media company, allowing social media companies to generate vast databases of peoples' likes, conversations, online behaviour, personal information, etc. all of which become invaluable to advertisers and others with vested interests in knowing how people think and feel. Each of Facebook, Twitter, Tumblr, Instagram,

and YouTube serve a somewhat different purpose, operate in different ways, and have control over different segments of the market. Each have unique regulatory challenges and strategies, though there are also relevant areas of overlap and agreement.

The Facebook landing page states their purpose: “Facebook helps you connect and share with the people in your life” (*Facebook*, 2019). It is a platform primarily intended for individuals to communicate with people in their personal networks. Twitter’s homepage provides a similarly pithy mission statement. “Follow your interests. Hear what people are talking about. Join the conversation. See what’s happening in the world right now” (*Twitter*, 2019). While people do engage in personal interaction on Twitter, their primary purpose is to provide news, updates, and a sense of awareness of a broader conversation happening locally, nationally, or globally. Instead of a list of friends to receive updates from like Facebook, the Twitter follower model encourages users to follow and interact with strangers from shared backgrounds or with shared interests. Instagram simply offers one function: “see photos and videos from your friends” (*Instagram*, 2019). While Twitter uses threads, retweets, hashtags, and replies to foster the spread of dialogue, Instagram is just a feed of images. Users can like and comment on others’ images, but that is essentially it. YouTube combines the public, global focus of Twitter with the image-centred feed of Instagram, through the medium of video. Their “mission is to give everyone a voice and show them the world” (*YouTube*, 2019). Anyone can start a ‘channel’ and begin uploading videos of just about anything. The YouTube homepage shows the user a selection of these videos, based on subscriptions, previous watch behaviour, popular videos in their location, and a myriad of other factors unknown to us. Positioning YouTube with these other social media platforms might not be the most obvious choice, but I consider it relevant due to its commenting and subscription features that bring about interaction between video creators and viewers.

Using NVivo, I coded all the policies by the mode of Lessigian regulation they exemplified and performed a critical discourse analysis. For example, the phrase “we trust you to be responsible” (*Policies and safety*, 2019), from YouTube’s policy, is evoking regulation through norms and was coded as such. The research is concerned with examining different ways of using language that may have changed and identifying themes that have waxed and waned in relevance over time. Lessig’s four modes of regulation: norms, laws, architecture, and the market, serve as a container for the literature. With critical discourse analysis, we can see the machinations of regulation at work through the methods delineated by Lessig.

## 2.4 Literature Review: Themes and Focus

In the literature review chapters, I focus on the four main themes that are relevant to understanding Internet regulation. They are the law, content moderation, networked publics, and sex and sexuality online. The law chapter discusses major laws, primarily in the United States, that have shaped the online landscape. These include free speech law, pornography law, intellectual property, sex trafficking, and decency and obscenity laws. United States laws are the most relevant because these social media companies are headquartered, built, and managed there. United States laws and ethos inform their decision-making both from a liability perspective and one of cultural and moral viewpoints. Content moderation encompasses the practices undertaken by social media companies to regulate their users' content and behaviour. These practices are automated content detection and deletion, user flagging and report features, and human moderation. All of these are used in concert to maintain the platform. In the chapter on networked publics, I go in to the history of the sociological/political theory concept of the public, how it has been adapted to apply to Internet communities, and why it is useful to study social media as a networked public. When discussing sex and sexuality online, all these elements of the law, content moderation, and publics come into play. The relevance of these four themes will be evident, to each other and to the research, through analysis of the social media policies.

## 2.5 Lessig's Regulatory Modes

Lawrence Lessig's work fits well with critical discourse analysis because of his focus on regulation and control. His adage, "code is law" was ground-breaking in Internet studies because of how it shifted the onus of regulation from just lawmakers or the government to acknowledge how the writers of code were building the available space in which one could operate online. Lessig introduced the four modalities of Internet regulation in his 2000 book, *Code is Law*: norms, laws, architecture, and the market. I will define each one to be relevant to my purposes. Then I will use the four to structure my analysis and categorise the social media policies by which modes they are exemplifying. Sometimes the modes come into conflict with each other in ways that will be analytically useful.

Norms are the ways communities self-regulate, through established acceptable behaviours. These can be explicit or implicit, and can be inflicted externally and adopted by the community,

or built from the ground up within the community. Norms can change over time, and conflict results from not knowing or disregarding the norms, or when individual communities interact and try to reconcile their divergent norms. Laws regulate the Internet by threatening consequences for undesirable behaviour, on the part of Internet users and its architects. Architecture is all the ways the Internet is constructed, and the affordances of those design and code decisions. Its regulation can often be unseen, but architecture shapes the very spaces we enter online, let alone the conversations we have once we are there. The market inflicts regulation because social media companies exist as profit-driven ventures within capitalism. Anything that loses them money will probably not be around long.

## 2.6 Critical Discourse Analysis

Critical discourse analysis is the examination of the ways language reflects and reinforces social organisation. Quoting Fairclough and Wodak (1997), Wodak and Meyer state that critical discourse analysis views “language as social practice” (2001, p. 5) and “discourses as relatively stable uses of language serving the organising and structuring of social life” (2001, p. 6). When critically analysing discourse, one can use some key questions to interrogate the text, such as: “what assumptions is this text grounded on? What is included and excluded, privileged and silenced? What does the text do? What effects/actions does it achieve?” (Barker and Scheele, 2016, p. 201). This method is valid for examining social media policies because policies are ever-changing, multi-authored documents that seek to inflict control on a space. In the parlance of Lawrence Lessig, these policy texts can be understood as the norms that encourage and inform regulation. Norms, expressed through discourse, which promote an ideology. Fairclough considers an ideology to be a schema or worldview that is used to build hegemony (2015).

“Hegemony is the power over society as a whole of one of the fundamentally economically defined classes in alliance (as a bloc) with other social forces, but it is never achieved more than partially or temporarily, as an ‘unstable equilibrium’.

Hegemony is about constructing alliances and integrating rather than simply dominating subordinate classes, through concessions or through ideological means to win their consent. Hegemony is a focus of constant struggle around points of greatest instability between classes and blocs, to construct or sustain or fracture alliances and relations of domination/subordination, which takes economic, political and ideological forms” (Fairclough, 1995, p. 76).

Here, Fairclough is drawing on Foucault’s work on power relations.

Critical discourse analysis will be used to examine if and how adult content policies (re)produce societal restrictions on women, LGBTQIA people, people of colour, and sex workers. By controlling the ways in which marginalised people can speak online, their marginalisation is maintained. As Norman Fairclough states; “power in discourse is to do with powerful participants *controlling and constraining the contributions of non-powerful participants*” (2015, p.75, emphasis his). Critical discourse analysts usually take their understanding of power from Foucault: power is structural and executed on a systemic level (Wodak and Meyer, 2001). But wherever there is power there is also resistance; power relations are about the struggle to inflict or resist power (Barker and Scheele, 2016). Habermas (1991) also acknowledged the role of language in the construction of the public sphere and power relations. We will now examine language more closely as we consider the laws that have jurisdiction over online expression.

### **3. Literature Review: The Law**

Early Internet law especially was based on extant broadcasting and obscenity laws that regulated radio, television, and publishing. However, the translation was not fully successful, as the First Amendment to the US Constitution complicated regulation. ‘Obscenity’ is not protected by the First Amendment, but it is up to the courts to determine whether material is obscene in each individual case (Axelrod-Contrada, 2007). When these judgments must be made on the scale of all content distributed online, it becomes untenable. The obscenity of material is determined by the Miller Test, named for the 1972 US Supreme Court case *Miller v. California*. The Miller Test asks

“(a) whether the average person, applying contemporary community standards” would find that the work, taken as a whole, appeals to the prurient interest... (b) whether the work depicts or describes, in a patently offensive way, sexual conduct specifically defined by the applicable state law, and (c) whether the work, taken as a whole, lacks serious literary, artistic, political, or scientific value” (Burger, 1972).

#### **3.1 The Communications Decency Act**

The Communications Decency Act was originally an amendment to the Telecommunications Act of 1996, intended to protect children from encountering obscene or indecent material online. The CDA criminalised knowingly sharing or transmitting indecent material with someone under

the age of 18. The Communications Decency Act was mostly struck down by the US Supreme Court in 1997, in the case of *Reno v. American Civil Liberties Union*. The court found that the case for protecting children from harm was not strong enough to infringe on the right of adults to free speech, and removed the criminal liability element from the CDA (Kuzma, 2013). While certain categories of speech--obscenity and child pornography--are not protected by the First Amendment, other things also under the nebulous umbrella of 'indecent' are protected speech (Godwin, 2003). While obscenity and child pornography have codified legal definitions, Smith argues that the notion of what constitutes 'harm' is fluid and has varied wildly over time and by culture (2000). Quoting literary scholar Walter Kendrick's 1987 tome on the history of pornography, *The Secret Museum*, she notes that the concept of the inherent vulnerability of young people came to be in the Victorian era, and "dwells exclusively in the mind of the archbourgeois" (cited in Smith, 2000, p.389). In *Reno v. ACLU* (1997), the Supreme Court stated that overly restrictive suppression of content aimed at adults in the interest of protecting children was unconstitutional; the adult's right to protected speech should take priority (Smith, 2000). An electronic magazine for gay and lesbian teenagers was one of the plaintiffs in this case; the teenage contributors were concerned their writing would be considered indecent because it dealt with sexuality (Axelrod-Contrada, 2007). *Reno v. ACLU* became a watershed case for free speech online in that it generally discredited the 'think of the children' defence. What remained of the CDA after *Reno v. ACLU* is a provision stating that "no provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider" (Kuzma, 2013, p. 33) and that the provider cannot be held liable for either what someone else posts on their service or for any action the provider chooses to take to limit or restrict objectionable posts. Essentially, the website provider can allow or ban 'indecent' content at their own discretion.

As Kuzma notes, the CDA does take a different stance than the Digital Millennium Copyright Act of 1998, where service providers are not entirely immune from liability, but are protected under certain conditions: "(1) it must adopt and reasonably implement a policy of terminating in appropriate circumstances the accounts of subscribers who are repeat infringers; and (2) it must accommodate and not interfere with 'standard technical measures'" (US Copyright Office, 1998, pp. 9-10).

These liability limitations were first tested in the case of Napster, a peer-to-peer file sharing service that allowed the widespread sharing of copyrighted music. The court found that Napster



had knowledge of the copyright infringement and did not take appropriate steps to stop it, therefore making them liable (Barako, 2001). So in reality, the DMCA makes it fairly easy to stop copyright infringement, as long as one has a legal team ready to challenge the infringer in court. A later case involving YouTube established immunity from secondary liability as long as the host complied with DMCA takedown notices (Charlesworth, 2011). These two cases show how the safe harbour defence can be unequally applied when it comes to the liability of internet service providers. While copyright and intellectual property are not the direct focus here, they have long been at the forefront of the discussion around Internet freedom and regulation. “As control over ideas and knowledge has become a key economic resource and source of power, it affects the content available on social media and the very use of these platforms” (Hintz, 2016, p. 331). It is worth considering the implications of intellectual property online, especially when billions of people are generating content every minute on these corporate platforms.

### **3.2 Child Online Protection Act**

The Child Online Protection Act (COPA) of 1998 was Congress’s response to the failure of the CDA. It was crafted to be narrower in scope to avoid constitutional challenges, but to accomplish the same goal of restricting explicit content considered ‘harmful to minors’ and criminalising its purveyors (Roodenburg, 2001). As soon as it was passed in October 1998, however, COPA faced an immediate lawsuit from the ACLU. The first temporary restraining order was issued in November of that year. The court found that COPA was invalid for the same reason as the CDA: there was no way to block minors from accessing harmful or indecent material without also preventing adults from exercising their right to free speech. The *Reno v. ACLU* decision was key precedent (Roodenburg, 2001). After several injunctions, COPA was finally struck down entirely by the Supreme Court in 2009, in *ACLU v. Mukasey*. Several key flaws in COPA are relevant for later Internet legislative attempts. One, adult age verification barriers such as requiring a credit card “deter adults from accessing protected speech because they impose costs on content that would be free, eliminate privacy, and stigmatize content” (*ACLU v Mukasey...*, 2009). Two, by using ‘community standards’ to determine what is harmful to minors, “COPA essentially requires that every Web publisher subject to the statute abide by the most restrictive and conservative state’s community standard in order to avoid criminal liability” (Roodenburg, 2001, p. 242).

### **3.3 Stop Enabling Sex Traffickers Act**

Some legal scholars are pleased with the 2018 passage of the Stop Enabling Sex Traffickers Act and the Allow States and Victims to Fight Online Sex Trafficking Act (collectively known as SESTA-FOSTA), saying that it closes a longstanding loophole: Section 230 of the CDA, sometimes known as the safe harbour clause (Leary, 2018). It is known as a safe harbour clause because it protects Internet Service Providers (ISPs) and ‘interactive computer service providers’ from liability for speech posted by a third party but hosted on their service. Other countries have no such laws, making the United States a ‘safe harbour’ from which service providers can operate. Section 230 has allowed social media to flourish. YouTube can host millions of videos without legal repercussions for what one of them might contain, millions of people can post on Facebook every day without Facebook facing charges when some of the posting is illegal content. SESTA-FOSTA changes this, at least for service providers who are “knowingly assisting, supporting, or facilitating” sex trafficking, sex by force or coercion, and unlawful prostitution. SESTA-FOSTA was supported by other groups who wish to increase the power of the state to criminalise speech online, who see this first exception to Section 230 as a crack to weaken free speech. These groups include 21st Century Fox who perhaps are looking towards a future where piracy is another amendment to Section 230, and tech giant Oracle, who were apparently backing the bill out of animosity towards Google (Johnson, 2017; Jeong, 2017).

Others fear that weakening the protections towards free and uncensored speech online in any way will inevitably lead to further restrictions once the precedent is set. Hintz examines how countries have treated “social media as an object of policing” (2016, p. 328). Perhaps the most ambitious example is ‘The Great Firewall of China’, but filtering or blocking content on religious, moral, or political grounds has become widespread, especially in the Middle East and Asia. SESTA-FOSTA is an example of how moral attitudes are spreading into Internet legislation in the ‘Western world’.

### **3.4 Sex Worker Response to SESTA-FOSTA and the Rescue Industry**

The rescue industry is a term coined by anthropologist Laura Agustín to describe the lucrative nonprofit industry of ‘rescuing’ women from sex work (Grant, 2014). The rescue industry “derives value from the production of awareness: It gives the producers jobs ...Raising

awareness serves to build value for the raisers, not for those who are the subjects of the awareness” (Grant, 2014, p. 37). Lawmakers who support sex work prohibition laws produce moral standing: tough on crime, tough on borders, ‘rescuing’ victims. “But even the most well-intentioned antitrafficking interventions have wrought significant ‘collateral damage’—much of it borne by the very groups most vulnerable to trafficking and exploitation” (Kinney, 2014, p. 147).

Many actual sex workers and sex worker advocate groups are opposed to SESTA-FOSTA, pointing out that the now-illegal platforms for screening clients online makes sex work safer (Mac and Smith, 2018). In a survey of over 600 UK-based sex workers, 75.2% said the Internet was a ‘very’ or ‘quite’ important resource for their workplace safety (Sanders et al, 2017). The broad language of the bill is such that law enforcement can exercise a lot of discretion on which websites to target. Opponents’ contention is that by conflating sex trafficking and sex work, “framing prostitution as de facto commercial sexual exploitation and overstating the evidence on its overlap with trafficking” (Cockbain and Bowers, 2019) Congress has failed to understand the landscape they are regulating (Renegade and Pottenger, 2019). Essentially, there is one key disagreement. Proponents point to escort advertising websites (Backpage, etc.) as loci for sex traffickers--where pimps advertise their victims’ services (Kuzma, 2013; Leary, 2018). However, according to the sex workers who participated in Sanders et al’s 2017 study, the Internet has in fact overwhelmingly enabled them to “work independently without having to rely on third parties”, decide where and when to work, “find out about their rights at work”, and helped them build peer support networks (p. 2). 78.3% responded that the Internet has improved the quality of their working lives (Sanders et al, 2017).

Legislators tend to overweight sexual exploitation of women and girls against other types of human trafficking, such as labour or domestic service trafficking (Cockbain and Bowers, 2019). Historical links have been drawn between this tendency to sexualise and gender anti-trafficking efforts and the ‘white slave trade’ panic of the early 20th century--“entwined with an abolitionist stance on prostitution, a moralising agenda, xenophobia and fears of racial contamination” (Cockbain and Bowers, 2019, p. 2). Additionally, sex trafficking policy in the United States and United Kingdom has been influenced by the “moral crusade” led by the religious right and some radical feminists against pornography and sex work (Weitzer, 2007). Weitzer defines moral crusades as “movements [that] define a particular condition as an unqualified evil, and see their mission as a righteous enterprise whose goals are both symbolic (attempting to redraw or bolster normative boundaries or moral standards) and instrumental (providing relief to victims,

punishing evildoers)” (2007, p. 448). SESTA-FOSTA seems like a clear descendent of the anti-pornography moral crusade of the 1980s. In the 1990s, “research and activism by Thai women’s rights organisations...provided ammunition for Western feminists opposed to prostitution and the globalising sex industry...the icon of the trafficked, sexually exploited, ‘suffering third world prostitute’ was invoked repeatedly” (Kinney, 2014, p. 150).

Some say that silencing and marginalising sex workers was in fact the goal of SESTA-FOSTA, that it has been a “huge success” (Cooper, 2018) and “makes exploitation easier” (Mac and Smith, 2018, p. 124). Whatever its intent, its impact has been immediate: Craigslist Personals and Backpage.com shut down, and many platforms have changed their terms-of-service (*Community standards*, 2019; *Restricted content*, 2019). While SESTA-FOSTA is a US law, it affected the businesses of sex workers elsewhere too; “sex workers in the UK also had to scramble to move adverts onto different sites and servers, losing work in the meantime” (Mac and Smith, 2018, p. 125). boyd warns of these perhaps unintended consequences of political interests exerting their will on the Internet, saying “many regulatory moves intended to protect vulnerable populations backfire” (Jenkins, Ito and boyd, 2016, p. 118). Weitzer would more cynically say that the intent is not to protect vulnerable populations but to use their vulnerability as a cudgel to advance a political goal: “recounting the plight of highly traumatized victims is intended to alarm the public and policy makers and justify draconian solutions” (2007, p. 448). However, there is little meaningful support for actual victims of sex trafficking. A guidance released by the UK Home Office stated that female asylum-seekers from Nigeria “may be subject to reprisals or re-trafficking” (Home Office, 2019, p. 8) but goes on to caution that “trafficked women who return from Europe, wealthy from prostitution, enjoy high social-economic status” (Home Office, 2019, p. 8). This supposed wealth is grounds to deny them asylum--“it’s basically inviting Home Office decision-makers to refuse people forced into sex work” (McKinney, 2019). Charities who work with trafficking victims and asylum-seekers state they have seen no evidence that trafficked women become wealthy (McKinney, 2019). San Francisco Bay Area news outlet KPIX 5 reported that while most violent crimes decreased in 2018, human trafficking saw a 170% increase. They draw a connection between the shutdown of sites like Backpage and the dramatic increase in exploitative street-based sex work, as pimps used SESTA-FOSTA to place themselves back in power. Bay Area local harm reduction clinic St. James Infirmary estimated a tripling in street-based sex workers in 2018 (Steimle, 2019). Across the United States, sex workers reported a dramatic increase in contact and threats from former pimps and managers. “As one sex worker said, ‘there’s always something in these

messages alluding to the bills that just passed. “*Now you need me.*” It’s really creepy because that exact thing is what the people who passed the bill thought they were fighting, and they’ve brought it into my life.” (Mac and Smith, 2018, p. 124). Cockbain and Bowers (2019) raised the additional concern that SESTA-FOSTA has, instead of stopping sex trafficking, has only pushed traffickers to darker and deeper corners of the Internet.

“Sex workers should not be expected to defend the existence of sex work in order to have the right to do it free from harm” (Grant, 2014, p. 39). Just released in July 2019, the new UK Conservative Party Human Rights Commission report evidences how little sex workers are listened to when policies are being written about them (Dorman, 2019). “The Commission supports the principle of evidence-based policy-making, but legislative approaches surrounding prostitution ought not to be determined by weighing the available evidence in isolation” (Dorman, 2019). Instead of using the available evidence that shows criminalising sex work only serves to further marginalise sex workers, that addressing poverty and facilitating legal migration does far more for sex workers’ well-being (Mac and Smith, 2018), the Conservative Party wishes to impose their moral values upon legislation.

Understanding Internet legislation is important because, as we will discuss next, the Internet has become a nexus for public life and social organising.

## **4. Literature Review: Networked Publics**

There is lots of research about the role of social media platforms as networked publics (boyd, 2010). Defining social networking sites as networked publics is important to show their relevance to and connectivity with life being carried out in the offline world. A networked public is simultaneously the virtual space created by and accessed through networked technologies and the “imagined collective that emerges as a result of the intersection of people, technology, and practice” (boyd 2010, p. 39). While in many ways a networked public serves the same function as the traditional idea of a public, the different structure that results from being mediated by technology forms new dynamics. As developing technologies “alter the societal architectures of visibility, access, and community, they also affect the contours of the public sphere, which in turn affects social norms and political structures” (Tufekci, 2017, p. 6). Technology has shifted the power dynamic of the public because of the relative ease of dissemination, communication, and access in the networked space. These new publics also

differ from the traditional ones in the ways information is conveyed and displayed: “according to a criteria of personal relevance... [and] conducted mainly in a conversational mode” (Schmidt, 2014, p. 4). Social media breaks down the barriers between the speaker and the listener and makes the connection explicit. Each individual user cultivates a personal context. “People turn to social networks to be comforted when they are feeling down..., to be built up when they are feeling insecure..., or, more generally, for positive emotional experiences.... Thus, as more and more people log in to these sites, affect accumulates” (Pybus, 2015, p. 237). The participant’s experience of the networked public is often guided by their affective response to information received, and then how they choose to share, respond to, or incorporate that information. Pybus defines affect as “a relational process that can propel bodies to act...in relation to social, economic, and political discourses” (2015, p. 240).

## 4.1 The Political Public

People use the public sphere to conduct discourse and form their opinions on public issues (Habermas, 1991). Fairclough adds that, from a Marxist perspective, the public sphere is also where ideology is disseminated, serving to maintain unequal power relations (2015). Using Foucault’s understanding of power, “these power relations are not stable but are constructed and reconstructed in discourse” (Neumayer and Valtysson, 2013, p. 9). Nancy Fraser redefined the public sphere through a feminist lens, complicating the original work of Habermas. She states that “public spheres are not only arenas for the formation of discursive opinion; in addition, they are arenas for the formation and enactment of social identities” (Fraser, 1990, p. 68). The link between identity and public speech is important, as Fraser notes that participation is not an equally afforded opportunity and true “participation requires being able to speak ‘in one’s own voice’” (1990, p. 69). Language can be both a tool for hegemonic control and a resistance against it. “Modern views of language as code and competence assume a unified and homogeneous social world in which language exists as a shared patrimony--as a device, precisely, for imagining community” (Pratt, 1991, p. 36). People form parts of their identities on the basis of their relations to others and perceived group membership. In the case of protest publics, their group membership is defined in opposition. When studying *networked* publics, the identity formation aspect is unmistakable: “profile generation [creating a social media profile] is an explicit act of writing oneself into being in a digital environment” (boyd, 2010, p. 43). Fraser (1990) also coined the term ‘counterpublics’ to refer to the phenomena of groups forming to speak up against and to counter the hegemonic narratives, as marginalised groups do not have

access to the classical public sphere. “Counterpublic address allows those that lie outside of sanctioned publics to map their own ideologies, thoughts, and subjectivities among people, mostly strangers, that share an awareness of similar countercultural referents” (Renninger, 2015, p. 1526). Renninger studied the identified the affordances of social media that facilitated counterpublics forming online, including the persistence of ideas and information without necessarily being tied to a specific identity or user through replicability and searchability. Hashtagging has been another key component of networked counterpublics coalescing (2015).

## 4.2 The Cultural Public

Jenkins, Ito, and boyd (2016) say that a public can also be a group centred around a cultural product, producing and consuming media, though the Habermas-aligned political theorists disagree with the notion of a depoliticised public. However, the affordances of social media are such that all kinds of conversations are happening at all times in the same virtual space. Users of social media are simultaneously producers and consumers, and simultaneously engaged in political discourses, media fandoms, interpersonal interactions, and more. When looking at Twitter, Bruns and Moe identified three layers of communication that interact and inform each other to create the public-at-large: “the micro level of interpersonal communication, the meso level of follower-followee networks, and the macro level of hashtag-based exchanges” (2014, p. 16). While a profile is technically public, users tend to not view their audience as the entire population of the Internet. Instead, “they imagine an audience that is usually more constrained by who they wish to reach and how they wish to present themselves” (boyd, 2010, p. 44). There is both a literal public-at-large and the imagined or intended public when speaking online. Recommendation algorithms also help build the sense of the imagined collective on social media by showing users other people with similar interests or commercial demographics (Gerrard, 2018). The imagined collective also has imagined norms, where users come into conflict over a set of implicit behavioural standards. “Possession of this implicit knowledge about shared routines and expectations become a condition of inclusion” (Schmidt, 2014, p. 7). So while there is no consensus on what a public is, if there is one or many, around what they organise, and to what end, there are some general themes we will be working with. Current thinking generally agrees that there is no unified public, but rather many publics and counterpublics where different ideological groups gather to create a unified sense of identity. The public sphere is at the intersection of politics, ideology, culture, and identity. It is the nexus

of discourse. “The thriving of publics, which are constantly morphing, relies on information moulded in counterpublics to push publics into new directions” (Renninger, 2015, p. 1526).

The idea of multiple publics in conversation with one another is similar to Mary Pratt’s theory of contact zones. She defined contact zones as “social spaces where cultures meet, clash, and grapple with each other, often in contexts of highly asymmetrical relations of power, such as colonialism, slavery, or their aftermaths as they are lived out in many parts of the world today” (Pratt, 1991, p. 33). Today, this theory is applied across many disciplines in the social sciences, including postcolonial theory and critical race theory. Additionally, there is a conceptual link between contact zones, publics, and the Foucauldian envisioning of the archive as “an important space of interpretation and contestation that has the power to make meaning through its ability to privilege certain discourses over others” (Pybus, 2015, p. 239). In all of these concepts there are groups with unequal power relations which produces tension, and differing valuations of discourses that uphold and challenge power. Archives are another pillar of the imagined collective, in that they are the repositories for stories that people engage with to build their identity within a group (Cho, 2015). “Social problems are couched in public and private discourses that shape the definition of these problems as well as inhibit productive social change” (Wodak and Meyer, 2001, p. 140).

### **4.3 Activist Counterpublics**

Tufekci (2017), Neumayer and Valtysson (2013), and others have studied how social media and Twitter specifically serve as a venue for protest and dissent, as a networked public that coalesces around political issues. However, “[social media] can also be considered oppressive due to factors such as ownership, use and control of information, surveillance, and privacy” (Neumayer and Valtysson, 2013, p. 4). This is an interesting theme that arises often in discussions of networked publics and social media in general, that they can be both liberating and oppressive. Counterpublics arise to discuss their lived issues and social media provides an avenue to broader awareness that had been denied them previously. But they are also subjected to censorship and surveillance, both from the state and private individuals who have a vested interest in maintaining hegemony. Networked publics coalesce suddenly, defined by oppositionality, with communication flowing between platforms. “There is a new, radically different mode of information and attention flow: the chaotic world of the digitally networked public sphere (or spheres) where ordinary citizens or activists can generate ideas, document



and spread news of events, and respond to mass media” (Tufekci, 2017, p. 29). Anyone with a smartphone can broadcast live video and people can communicate instantaneously across the world. Tufekci identifies the ways authorities have also adapted to technology’s role in social movements with their own methods of control: “producing information glut, inducing confusion and distraction, and mobilizing counter-movements, rather than imposing outright censorship, are becoming parts of the playbook of governments that confront social movements” (2017, p. 7)—though of course outright censorship is still widely used by oppressive state forces.

## 4.4 Lessig’s Regulation and the Public Sphere

When the public is networked and ‘code is law’, the architecture of the space shapes the available and permissible discourse (Lessig 2000). “Similar to the replacement [of the town square by the mall], the privatised infrastructure of commercial social media offers a confined and controlled space for semi-public interactions, under the conditions of a commercial logic” (Hintz, 2016, p. 327). boyd identifies four elements of social media sites that facilitate and construct their role in networked publics: profiles, lists of friends/followers, feeds/timelines of updates, and public commenting tools--usually in addition to private messaging avenues (2010). The timeline helps create the public because it gives people a real-time sense of the activities of their fellow users. Profiles do so because it is the way a person establishes their presence in the network. Friend lists allow people to delineate their online communities and public commenting allows others to communicate in the public sphere, while there is also a simultaneous venue for private interpersonal communication on the same platform. Changing modes and standards of online content moderation change the nature of the public and enforces hegemony. The architecture of networked publics is a key difference from the conventional public due to the ability to control and shape the digital environment, and through the power to influence and restrict the discourse occurring in the digital environment. “Technologies reproduce existing hierarchies and systems of power but can be appropriated for subversive action to challenge power and lead to political change” (Neumayer and Valtysson, 2013, p. 4).

Lessig’s view on the role of the market is also visible when considering the character of the networked public. “The media that constitute the material support for the circulation of views are privately owned and operated for profit. Consequently, subordinated social groups usually lack equal access to the material means of equal participation” (Fraser, 1990, p. 64). Companies espouse ideals of freedom, sharing, and collaboration but “underneath the rhetoric of

participation are mechanisms of regulation and control” (Jenkins, Ito, and boyd, 2016, p. 102), due to the need to maintain profitability and appease venture capitalists and stakeholders. Neumayer and Valtysson noticed that despite the “apparently flat hierarchies” (2013, p. 12) of social media, some users, including journalists, can become central figures important to disseminating information. Tufekci agrees, stating that the “result is a new public sphere that is more open than in the past, but one that is not flat in the sense of all information and nodes having equal reach, attention, and credibility” (2017, p. 39). Attention itself has become the main currency of the current age, with those able to hold attention wielding the most power (Tufekci, 2017). Pybus (2015) adds that ‘stickiness’, or the amount of time spent on a particular website, is the most important metric of success for a social network. “Without people actively logging in and inputting data, such sites would quickly lose market value” (Pybus, 2015, p. 237). User-generated content is the foundation on which social media companies create their value. However, the networked publics created by and accessed on social media provide their users with value as well, both in a political and affective sense.

## 5. Literature Review: Content Moderation

Content moderation occurs on a data scale but the platform’s users feel it on an intimate scale (Gillespie, 2018). There is a tension between the personal, human nature of social media and its vast structures of regulation and surveillance. In his 2018 book *Custodians of the Internet*, Tarleton Gillespie uses the case study of a Facebook group, “Hey Facebook, Breastfeeding Is Not Obscene!” to track user activism that eventually resulted in policy change. He also notes the ways in which the complete opacity of platforms’ architecture and decision-making are harmful to users when they run afoul of unknown policies. Platform moderation is an “internalist concern--what should the platform’s rules be, how will they be enforced, what is a fair way to govern users?--and an externalist concern: how does what the platform allows or prohibits affect broader public knowledge and value?” (Gillespie, 2018, p. 179).

The case of “Hey Facebook, Breastfeeding Is Not Obscene!” (Gillespie, 2018) reveals how context and intent cannot be separated from content moderation. This Facebook group was created to protest Facebook’s decision to delete images of women breastfeeding from the site. The female breast exists in many contexts, from sexual to medical to life-giving. The decisions made regarding in which context the breast is acceptable reveals a lot about Facebook’s values and those of the powerful it is attempting to appease. Lessig’s norms are evidenced here.

Facebook is following a certain set of norms, unchallenged before “Hey Facebook..!” that inform its decisions about the obscenity of the female breast. Eventually, after sustained protest, Facebook now allows breastfeeding photos where the breast is visible.

Difficulties in moderation are highlighted by cases where a bad faith actor uses an actual policy to their own political ends. In 2014, Twitter user/troll @RealNamePolice used Facebook’s Real Name Policy to get hundreds of drag queens banned or locked out of their accounts, motivated by homophobia (Gillespie, 2018). The Real Name Policy was just that—Facebook users must go by their legal name on their profiles. As the policy was written, @RealNamePolice was correct: those drag queens were not using their legal names on their Facebook profiles, which did violate Facebook’s Real Name Policy. It was only after widespread community outcry that Facebook addressed the issue by apologising, though not materially altering the policy. One of the affected drag queens described the Internet as a space “in which new forms of speech acts or gestures contribute to the production and disciplining of our subjectivities”, basically saying that it was important to be able to present as one’s authentic, queer self online, from a perspective of “theories of gender... as performative in a Butlerian sense” (Lil Miss Hot Mess, 2015, p. 145).

In 2015, an op-ed by Twitter general counsel Vitaya Gaddle in the *Washington Post* positioned moderation as key to allowing free speech to proliferate, because trolls, abusers, and harassers were so effective at shutting down conversations. “Freedom of expression means little as our underlying philosophy if we continue to allow voices to be silenced because they are afraid to speak up...protection from threats and abuse will allow users to remain and opinions to flourish, expanding diversity and debate” (quoted in Gillespie, 2018, p. 31). Yet by its very nature, some moderation is silencing valid speech, with real world consequences. In 2014, reports emerged that journalists and organisations who were too critical of the Vietnamese government repeatedly had their Facebook pages deleted. The journalists’ belief was that Vietnam’s ‘opinion shapers’ (people employed by the propaganda office to influence online discussion) were mass-reporting their pages for abuse. The Facebook accounts of more than forty Vietnamese journalists were deleted (Brandon, 2014). Similar issues arose that same year for members of the Syrian opposition. “Activists believe groups supportive of Syrian President Bashar al-Assad are gaming the system and reporting on their rivals. Facebook does not disclose information about who reported whom, making it impossible to confirm these theories. But the pro-Assad Syrian Electronic Army (SEA)...has publicly gloated about this tactic” (Pizzi, 2014). The SEA

took credit for the NGO Syrian Network for Human Rights losing its Facebook page. As Pizzi pointed out, people rely on groups like SNHR for information on the atrocities committed by pro-Assad militants. Without their Facebook page, and its 42,000 like at time of deletion, the SEA has successfully limited their reach.

“Direct intervention by state authorities is increasingly complemented by the application of pressure on social media companies to police themselves” (Hintz, 2016, p. 329). Internal policies leaked to *ProPublica* revealed that at one time, Facebook “banned posts that praise the use of ‘violence to resist occupation of an internationally recognized state’” (Angwin, 2017). Therefore, Facebook had deleted posts from journalists and activists in regions such as Palestine, Kashmir, and the Crimea. “Platforms of this size tend to collapse contexts in favour of establishing global rules that make little sense when applied to content across vastly different cultural and political contexts around the world. This can, at times, have significant negative impact on marginalized groups....[Angwin’s] work demonstrated that attempts at universal anti-discrimination rules too often do not account for power differences along racial and gender lines” (Caplan, 2018, p. 25). This context-blind approach favours hegemonic control. When the state relies on businesses to implement and enforce communications policy, they “delegate responsibility to the private sector” (Hintz, 2016, p. 334).

## 5.1 Flagging and Reporting

The flagging/reporting method of content moderation provides a smokescreen of legitimate response to platforms. If they say, there is too much content for us to screen, we only look at flagged material, then they are not responsible for offensive content that proliferates. It also grants the appearance that the user base has a stake in the direction of the platform and some control over their experience, “that the platform is listening to its users and providing avenues for them to express offense or seek help when they’re being harmed” (Gillespie, 2018, p. 93). In reality, platforms can still choose to apply their policies however they see fit. A flag by a user does not mandate a response from the platform and this opacity grants them plausible deniability. The vague nature of the public policies is intentional, as evidenced by the separate, private documents used by the internal moderators that delineate much more clearly what constitutes a violation (West, 2018).

Relying on user flags is using Lessig's 'norms' method of regulation. Just about all social media platforms offer a flagging system. Respondents to West's 2016 survey mostly identifying other users flagging their content as the perceived agent of regulation (West, 2018). The policies help delineate the norms and then the flagging system invites the users to help police them. The language and the process all belong to the platform, so the platform can control expression of objection totally. The language of *community* standards or guidelines when is deceptive because it implies a singular community all together on the platform operating under the same norms. While in the literal sense of the word a community, social media platforms are hosts to many communities, who often operate at cross-purposes and have entirely different norms. "The term platform itself can serve as an obfuscating device, operationalized at times by social media companies in order to divert tensions between their obligations to multiple constituencies" (West, 2018, p. 4367).

Hintz uses a similar framework to the work of Lessig to examine how commercial policy, infrastructure, and state control intersect while regulating social media (Hintz, 2016). His conclusion is that social media companies "play an active role in developing and enforcing new rules for allowing as well as restricting information; they define and punish objectionable user behaviour; and they provide and withdraw, accordingly, vital spaces and resources for communication" (Hintz, 2016, p. 336). Social networks are an important space for non-hegemonic communication, but they are increasingly being subjected to censorship, filtering, and state surveillance--displaying two of Lessig's pillars of architecture and the law. "It is a curious fact of the participatory web that social media platforms sometimes find themselves housing their own critics" (Gillespie, 2018, p. 166).

## **5.2 Human Moderation versus Automatic Content Detection**

Content moderation has been in the news lately (Newton, 2019a, 2019b; Vengattil and Dave, 2019) as outsourced Facebook moderators complain of low pay and traumatising conditions. Whenever a post gets reported for violating one of Facebook or Instagram's Community Standards, it has to be reviewed by a real person. There are thousands of people worldwide who work for these contractors. They are paid dramatically less than the official Facebook employees in Silicon Valley and receive cursory mental health assistance while being constantly exposed to snuff films, the recruiting materials of violent extremists, conspiracy theories, and yes, pornography (Newton, 2019a, 2019b). Employees at one such contractor, Cognizant,

reported being under constant stress to meet Facebook's 98% accuracy marker and under omnipresent threat of firing for falling below it (Newton, 2019b; O'Connell, 2019). Turnover rates are not available, but all those interviewed said that no one made it longer than two years (Newton, 2019b). Moderators suffer from post-traumatic stress from the graphic violence they review all day and one said that even months after being laid off he still suffers "flashbacks, sleep disturbances, and feelings of disconnection from friends and family" (O'Connell, 2019).

Lessig's regulation through code is present in the increasing use of algorithms, artificial intelligence, and machine learning in content moderation. AI is an attractive option because many believe it solves the scale problem, removes the potential for harm to human moderators, and the (erroneous) belief that an algorithm can be a neutral arbiter who follows the letter of the law without bias or political motivation (Angwin, 2017). The reality is that "even automated tools and complex algorithms are designed and tested by people, enacted and maintained by people, and deployed or overridden by people" (Gillespie, 2018, p. 104) and are therefore influenced by human biases and judgment. The regulation is coded into the architecture but is still reliant on human decision-making, just shifting the responsibility from individual responses to offensive content to an examination of trends and patterns.

### **5.3 Free Speech Tensions**

Some Internet scholars advocate for more legislative oversight on content moderation, with the goal of preserving US ideals of free speech enshrined in the First Amendment (Langvardt, 2018)--or with a goal of mitigating widespread, effective disinformation and propaganda campaigns. However, the same attitude that feels Philando Castile's girlfriend should have been able to live-stream his murder at the hands of police without Facebook's intervention might also feel the Christchurch, New Zealand terrorist should have been able to live-stream his massacre. To some extent, this is the tension that has always existed at the heart of the United States' free speech. "The possibility of anonymous speech on the Internet, combined with the ease of 'one to many' communications, largely removes the normative and practical constraints that made content-shock rare in the twentieth century" (Langvardt, 2018, p. 1361). The difference now is the sheer volume of content flooding onto social media every second. Free speech defenders are advocating for Lessig's pillar of the law to step in to regulate the regulation. Langvardt (2018) uses some Supreme Court precedent to establish a basis for treating a company like Twitter as a state actor. *Marsh v. Alabama* decided that company-owned towns were to be

treated as “quasi-municipalities” and still bound to uphold First Amendment rights, a decision that was enforced in the case of privately-owned shopping malls in *Amalgamated Food Employees Union Local 590 v. Logan Valley Plaza, Inc.* (though *Logan Valley* has since been overturned). The Court generally holds certain types of private property as practical equivalents to public land, which allows for common-sense application of the First Amendment. It is a stretch to get from a company town, which is more literally comparable to a regular town, to the virtual company town of Twitter, but one can see how this precedent is relevant. Quoted in Hintz (2016), Mueller notes “the regulatory trend that constantly emerges...is a shift of the responsibility for monitoring and policing Internet conduct onto strategically positioned private sector intermediaries” (2010, p. 149).

Research shows that online communities are adept at adapting to regulation, by changing their language enough to avoid algorithmic flagging, and using disclaimers to distance themselves from the community being regulated. Gerrard (2018) explored how pro-eating disorder communities circumvent moderation by changing tagging practices, using coded language, and employing plausible deniability. When Instagram banned ‘pro-ana’ (pro-anorexia) hashtags like #thighgap, #thinspiration, and so on, pro-ana Instagram users switched to #thyghgapp and the more ambiguous #fitspiration (Gerrard, 2018). Social media policy, moderators, and researchers are always behind the curve because they first have to find a problematic community, study its communicative modes, and then impose regulations, while the community can react ad hoc to new regulations or re-form on another platform. Now, we turn to some of the most contentious Internet communities, to examine their members and motivations.

## **6. Literature Review: Sex and Sexuality Online**

### **6.1 Porn: For Who, By Who?**

There is no academic consensus on the place of pornography in society. Pornography is both a venue for self-expression and for fetishisation and exploitation (Miller-Young, 2010). While some women may wish to post topless photos, others are being victimised by ‘revenge porn’. In feminist scholarship, there has been a long-ongoing battle between those who view pornography as violence, as it “endorses and encourages men’s abusive sexual desires and behaviours towards women” (Miller-Young, 2010, p. 222), and those who worry pornography is

an easy target for those who wish to regulate and control women's behaviour. Despite its positioning within heteronormative capitalism, pornography can still offer "spaces and possibilities for counter-appropriation, self-expression, pleasure, and labour" (Miller-Young, 2010, p. 222). Heteronormative capitalism refers to the way capitalism relies on heteronormative technologies of power to enable "subjects and bodies to be constituted, governed, normalized, and rendered 'natural facts' that, as such, play a key role in the (re-)production of capitalist society" (Ludwig, 2018, p. 136). Essentially, heteronormativity produces the social order and consensus that capitalism needs to function, while capitalism inflicts its logic of labour and production onto relational and family structures. Heteronormativity and capitalism produce and uphold each other.

Coletto et al (2017) studied pornography consumption as a 'deviant community' in the broader context of the social network at large. They found that the majority of pornography was being produced by a small core group of users but spreads widely across the network, "potentially reaching a large audience of people who might see that type of content unwillingly" (Coletto et al 2017, p. 28). This highlights some of the important context for policy-making. Unwilling exposure to pornography is a common hazard of Internet use that social media companies seek to mitigate. Coletto et al also delve into a broad demographic survey of adult content producers and consumers by age and gender. However, they specify their 'deviant networks' are those that revolve around hardcore pornography, not less explicit content such as "soft porn, artistic nudity, manga" (Coletto et al, 2017, p.28). Mainstream pornography is often criticised for reinforcing racist and sexist stereotypes and for its treatment of performers, particularly marginalised ones. "The hegemony of whiteness in defining human capital (including erotic value) disadvantages people of colour. The black female body is, in relation to other racialized and gendered bodies in the flourishing sex industries, devalued and marginalized in pornography's hierarchal, sexual marketplace" (Miller-Young, 2010). Black women's labour is devalued in this system, yet they also seek to create their own sexual representations. This is why online spaces outside of the mainstream have been so important for the sexual autonomy of marginalised groups. "We aren't asking you to love the sex industry. We certainly don't. We are asking that your disgust with the sex industry...doesn't overtake your ability to empathise with people who sell sex" (Mac and Smith, 2018, p. 208).



## 6.2 Feminist Pornographies

Feminist scholars have discussed the culture of pornography and erotica on Tumblr and other blogging sites (Mondin, 2017; Muise, 2011; Wood, 2008). Mondin says that “these archives of feelings aim at countering mainstream representations” (2017, p. 288) and describes the community of feminist and queer pornographers and viewers as “an online counterpublic that has found on the platform [Tumblr] easiness and freedom of access” (2017, p. 290). The archive is sociologically important because every archive privileges certain narratives, or discourses. Cho describes some elements of the ‘archives of feelings’ that allow queer people to share history and community due to their underground nature; they are “ephemeral, unofficial, evasive” (2015, p. 47). Muise analyses the ways in which sex blogging can challenge heteronormative structures and allow women to express their desires (2011). “A sex blog allows people to type themselves into being in a more open and layered way than, perhaps, some other form of non-sexualised social media participation” (Tiidenberg, 2013, p. 179). Wood draws a conceptual link between women’s sex blogs and the traditions of women sharing sexuality information, through oral traditions and in-person social networks. According to Wood, building a ‘sex commons’ online is “important feminist work, allowing women, individually and collectively, to reassert control over knowledge about their sexuality” (2008, p. 484). The ‘sex commons’ concept also includes discussion of laws regarding copyright and corporate control of information. In concert with Muise (2011), Wood also highlights the importance of sex blogs for destigmatizing desire (2008). Participants in Mondin’s survey discussed how their specific desires were not catered to by mainstream pornography, or how mainstream pornography could be off-putting due to its production for the ‘male gaze’ and its tendency to be “gross, totally uninteresting and stereotypical” (2017, p. 287). For respondents, pornography seen as politically or ethically questionable could be actively un-erotic. The experience of browsing and curating these ‘archives of feelings’ led to affective responses across the emotional spectrum (Cho, 2015; Mondin, 2017).

Cultural pressure forced Facebook and Instagram to relax their nudity policies incrementally: from a full breast ban to allowing “some images of female breasts that include the nipple...including those depicting acts of protest, women actively engaged in breast-feeding, and photos of post-mastectomy scarring” (*Community standards*, 2019). The long-running debate over women’s bodies on Facebook is documented in Gillespie’s *Custodians of the Internet* (2018). Gillespie notes how “tracking this dispute [regarding images of breastfeeding] with

Facebook highlights the impact platform moderation can have when it impresses itself on the complex terrain of social practice and meaning” (p.151). Demone (2015) complicates the nipple issue, noting that the right to bare breasts has only been granted in limited contexts, to a small set of relatively privileged bodies.

### **6.3 Networked Queerness**

Social media in general, and Tumblr specifically, are also central locations for exploration and declaration of sexuality, gender, and identity. “This is in step with a history of queer people using the Internet from its earliest iterations to express identity and articulate connection in a way that spans real and virtual geographies” (Cho, 2015, p. 46). Renninger (2015) identifies Tumblr’s affordances that make it key for counterpublic communication, including options for privacy/anonymity and ephemerality afforded by its unique tagging system that often functions as semi-private commentary. “The politics of authenticity are mobilised here against the perceived fakeness of the mainstream” (Mondin, 2017, p. 288). These themes of authenticity and identity performance and formation are echoed by Lil Miss Hot Mess (2015) in my Content Moderation chapter and Fraser (1990) in my Publics chapter. Oakley (2016) complicates the ‘authenticity’ discussion by delineating between an authenticity that relies on external validation and the ‘queer realness’ or ‘true self’. “Sense of self and the identity work one must do to get to their true self is a mediated venture that requires effort to fold one’s queer identity into their own ‘realness,’ rather than relying on an outside source for validation of authenticity” (Oakley, 2016, p. 7). NSFW (not-safe-for-work or explicit content) blogs can allow users to explore sexuality within a community of others doing the same, but these blogs can also facilitate identity construction and community-building in nonsexual contexts (Oakley, 2016). Renninger notes that many different queer and LGBTQIA counterpublics use Tumblr “as a venue for in-group communication” (2015, p. 1520); his research regards asexual Tumblr communities. Tumblr users have been able to craft their own new, intersectional communities: “queerness moved from the periphery to the centre and demanded not only recognition but also comprehension” (Cavalcante, 2018, p. 12). However, Cavalcante also describes Tumblr as a ‘queer vortex’ where feeling can be intense and all-consuming but ultimately short-lived (2018). Finally, the fully queer potential of Tumblr is limited by its position as a subsidiary to a global corporation. This became evident when Tumblr banned all NSFW content in 2018 and a lot of the queer infrastructure came crashing down. Even before the full-on NSFW ban, there were several incidents where changes to the site, such as new ownership and incorporation, resulted in

temporary censorship of all LGBTQIA content (Cavalcante, 2018). The queer Internet migration has also been problematized by Grant, who ties it to the “broader trends towards sexual gentrification, as identified by author and long-time AIDS activist Sarah Schulman” (2014, p. 44). In response to gentrification of historically gay neighbourhoods and police raids on gay gathering places, Schulman noted that, “gay life is now expected to take place in private, by people who are white, upper class, and sexually discreet” (quoted in Grant, 2014, p. 44).

“In a Foucauldian sense, LGBTQIA bloggers construct discourses within existing hegemonic norms...although the LGBTQIA community asserts and makes space for their own genders and sexualities, they use the framework of gender that has been constructed by the hegemonic binary” (Oakley, 2016, p. 9). Generally, I will use the term ‘queer’ to refer to communities not centred around explicit declamation of identity and in the sense of consciously relating oneself to a group outside of the heteronormative hegemony. I use ‘LGBTQIA’ to refer to the non-straight sexual orientation umbrella, or when groups organise around a specific sexual orientation. In instances where a member of the community is referring to themselves or their community, I will use whichever term they do. Queer theorists use the work of Foucault and Butler to argue that organising around some element of identity is problematic because it works to enforce the hegemonic idea of static categories of identity and the discourse of validity through confession, through ‘coming out’ (Barker and Scheele, 2016).

Adult content is partially regulated on social media with algorithms that are trained to identify photos containing nudity. Key identifying features are large uninterrupted swathes of skin-coloured pixels in combination with facial recognition, photo captions or tags, metadata, and so on (Gillespie, 2018). Machine learning classification is better at operating at scale, but human content moderators are usually also used. In Tumblr’s case, the machine flags all the content it classifies as explicit, then the human moderators become involved to review the decision if the user appeals the flagging (*Updates to Tumblr’s community guidelines*, 2018). Problems with this method become evident, as the Tumblr algorithm once flagged dozens of patent application illustrations as nudity, including patent or litigation documents for swimming pool floats, shoes, lounge chairs, and snack food products (Burstein, 2018).

## 6.4 Sex Work Online

Sanders, et al (2017, 2018) have spent the last several years researching the landscape of online sex work. Online sex work is multimodal, taking place across different platforms, apps, and individual websites. There are advertising platforms and directories, webcam and text chat services, as well as 'off label' use of sales websites (Craigslist, Gumtree), social media, and dating and hook-up apps. There are also platforms where workers can sell pornography or other commissioned content like photosets, paying for web hosting and technical expertise. An affordance of online sex work is the flexibility (Sanders et al, 2017). The demographics of sex workers are indicative of this flexibility and fluidity: "sex work is often not always the primary mode of work and is undertaken on a part time or sporadic basis, with sex workers using the features of digital technology...to manage their time and labour" (Sanders et al, 2018, p. 156). Sex workers often credit technology for the high degree of control it allows them over their work and lives. "It is also possible for many people to try out sex work, organized online and conducted in private, without becoming a known prostitute" (Grant, 2014, p. 45). Jones (2015) also notes the ways in which a move to Internet-mediated sex work has benefitted workers, from increasing physical safety and decreasing risk through client screening or performing distance services, to achieving higher wages through advertising and diversifying revenue streams. Online workers have an easier time attracting desirable clientele that they would be unlikely to find 'on the stroll' and not relying on third-party management (pimps) increases their profit margins. While some take evidence of sex work becoming more profitable as a bad thing, Sanders et al (2017, 2018) and Jones (2015) take the position that less exploitation of labour is good. Additionally, the Internet reduces negative contact with police: "street-based sex workers represent 85 to 90 percent of arrests, yet they make up only 10-15 percent of the market" (Jones, 2015, p. 562). Because people of colour and LGBTQ individuals are more often subject to police violence and marginalisation, and over-represented in sex work (Mac and Smith, 2018), reducing potentially dangerous police contact is a benefit.

A final key benefit is how the Internet has allowed sex workers to organise and advocate for themselves while still protecting their offline anonymity (Sanders et al, 2018; Grant, 2014), though others say use of the Internet fragments in-person social networks and support systems important to sex worker well-being (Jones, 2015). One thing sex workers have in common with other marginalised groups is gathering on social media or other closed platforms to form virtual communities to combat their marginalisation and share resources while not being physically co-

located (Sanders et al, 2018; Grant, 2014). However, there are new risks associated with online sex work, including online harassment, doxing, or privacy concerns (Jones, 2015). Grant (2014) also notes that the affordances of the Internet for surveillance can be hazardous to sex workers advertising online.

“The online commercial sex revolution provides opportunities for work, empowerment, flexibility and freedom for some, but equally the concerns over precarity of employment, criminalisation, exploitation and lack of protection if sex workers become victims of crime. Current laws and continued structurally embedded stigmatisation of sex work means that many online sex workers remain invisible behind the screen, denied access to full labour rights, full citizenship and access to social justice, pushed increasingly by current laws to work behind the screen” (Sanders et al, 2018, p. 163).

## 6.5 Sex Workers Organizing Online

As has been discussed, there are many issues facing sex workers today. Sex workers are subjected to violence, discrimination, marginalisation, and criminalisation. Their businesses are being squeezed out of legitimate Internet spaces, and their stated needs are rarely considered when policies are being written about them. Most resources and focus are directed towards the most visible forms of sex work, that happening on the street and across international borders. Yet street-based sex workers and sex trafficking victims are in the minority of all sex workers (Feldman, 2014). “Contemporary sex workers’ rights activists are well aware of these inconsistencies and contend that they often occur because sex workers themselves are rarely included in the process of knowledge production about their work and industry” (Feldman, 2014, p. 243). In order to share their own stories and advocate for themselves, many activist groups have begun to organise online. Feldman examines the case study of the ground-breaking sex worker activist blog, *Bound, Not Gagged* (2014). Melissa Gira Grant, whose 2014 book *Playing the Whore* is cited throughout this dissertation, was one of the co-founders of *Bound, Not Gagged*, along with Stacy Swimme. Swimme authored the now famous sex worker activist rallying cry, “nothing about us, without us” (Feldman, 2014, p. 243). *Bound, Not Gagged* was created as a platform for sex workers to openly discuss their experiences and the issues affecting them. The Internet made this possible due to the ability to remain anonymous or pseudonymous. Sex workers networked with and supported each other. It was a way to circulate information on local events and hear news from activists nationwide. Blog contributors also analysed and critiqued media coverage and research regarding sex workers.

Grant (2014) tells her personal history of sex work activism being rooted in building online friendships with other sex workers who lived in her city. Their friendships were spurred into activism by the arrest of Shannon Williams, a Berkeley sex worker who was arrested in her apartment by a dozen police in riot gear who did not let her get dressed before dragging her handcuffed, in lingerie, out into the street (Grant, 2014). This event catalysed the founding of the Sex Worker Outreach Project (SWOP) and eventually Grant's contributions to *Bound, Not Gagged*. Today, while much sex worker activism is rooted in providing local, grassroots services and resources, most groups use social media extensively for the same reasons *Bound, Not Gagged* blogged. Sex workers can speak publically about issues and events from an anonymous space, protecting them from social and legal recriminations. Sex worker-led activism groups worldwide use the symbol of a red umbrella in their Internet branding to indicate a shared ethos. DecrimNY, "working to decriminalise, decarcerate, and destigmatise the sex trades in New York" (DecrimNY, 2019) organises in-person activism using Twitter as a medium for communication. Twitter allows DecrimNY to support sister decriminalisation groups by signal boosting their messages. "The most potent source of untapped power for sex workers is not sexual liberation, social rebellion, or even money, but solidarity" (Mac and Smith, 2018, p. 219).

## 7. Social Media Policy Analysis

The key policies I am examining are the current policies from Facebook, Instagram, Tumblr, Twitter, and YouTube, as well as highlighting some important changes from the last several years. The previous policies mentioned are Facebook's updates from October and December 2018 and Tumblr's updates from 2017 and December 2018. These policies have been transcribed in the appendix. In the cases where policies are changed, I have only included the changes.

Fairclough sees three interwoven elements of discourse: "social practice, discursive practice (text production, distribution, and consumption), and text" and the "analysis of a specific discourse calls for analysis in each of these three dimensions and their interrelations" (1995, p. 74). We can see each element at work in these social media regulatory discourses. Discursive practice is evident in the examples we have discussed where a company will respond to criticism about their policy. Regulated groups will oppose the text at hand, demanding change

after the text has served its purpose of reproducing power relations. So we can see how the texts themselves also perform a social function of control. Social practice dovetails with Lessig's norms.

## 7.1 Laws

The law is an agent of regulation because it holds the power of material consequence. Social media companies have changed certain policies in order to comply with new laws introduced over time, or to pre-empt potential liability. Some of the major laws that regulate adult content online include obscenity laws, laws against child sexual exploitation (CSE) and nonconsensual pornography, laws against prostitution, and most recently the Stop Enabling Sex Traffickers Act and the Allow States and Victims to Fight Online Sex Trafficking Act (SESTA-FOSTA). As Fairclough identifies, the very names of these acts are imbued with ideology (1995). The law also regulates indirectly, by regulating norms, architecture and the market. For example, "the law uses taxes to increase the market's constraint on certain behaviours and subsidies to reduce its constraint on others" (Lessig, 2000, p. 127). To evidence how laws regulate norms, Lessig points to laws around sex education in public schools. These laws are inflicting regulation on the choice of what norms to instil in children in school. Laws influence code or architecture as well. The Americans with Disabilities Act (ADA) is perhaps the best example to map onto the experience of cyberspace. The ADA requires public physical space to be built in certain ways, similar to how Internet law can regulate the parameters of virtual space. "In each of these examples, law functions in two very different ways. When its operation is direct, it tells individuals how to behave and threatens punishment if they deviate from that behaviour. When its operation is indirect, it modifies one of the other structures of constraint" (Lessig, 2000, p. 132).

Facebook reacted to SESTA-FOSTA by tightening their restrictions on "implicit sexual solicitation" (*Community standards*, 2018). SESTA-FOSTA were signed into law in April 2018. In October 2018, Facebook removed the language banning prostitution from the Sexual Exploitation of Adults section of their Community Standards and created a new section titled Sexual Solicitation, which expanded the language of nonpermissible statements. "We know sexual solicitation isn't necessarily exploitative, and that the language we use...matters a lot" (*Community standards*, 2018). Facebook directly admits the power of their words here, doing the discourse analysis work for me. It is true that not all solicitation is exploitation, and

linguistically making that distinction has broader societal effect. However, these acknowledgements ring a little hollow, as all this distinction truly serves is banning all kinds of solicitation on Facebook. They use the language of sex work activism, but only to disallow it. Facebook did not acknowledge SESTA-FOSTA as a precipitating factor but instead obliquely refers to “some audiences within our global community that may be sensitive to this type of content” (*Community standards*, 2018). The use of “some people in our community may be sensitive to this type of content” (*Community standards*, 2019) is repeated in slightly varying ways across Facebook’s policy documents. Now, not only is explicitly advertising sexual services banned, but a wide array of oblique statements are also not permitted. “Vague suggestive statements, such as ‘looking for a good time tonight’” (*Community standards*, 2018) are not allowed. This literally vague language gives Facebook a wide margin in which to act. Banning ‘implicit sexual solicitation’ means that Facebook is taking it upon itself to parse intent from posts. Is ‘implicit sexual solicitation’ meaningful distinct from flirting? This Facebook policy could be weaponized by trolls, since otherwise innocuous discussions of sexual health or safety might be considered “knowingly assisting, supporting, or facilitating...unlawful prostitution” (Cooper, 2018) and Facebook might be held liable. We saw this play out already with the Real Name Policy and @RealNamePolice’s crusade against drag queens (Gillespie, 2018). Any kind of suggestive language, when reported enough times, can be grounds for deletion. Politically motivated groups already organise apparently successful mass report brigades (Brandon, 2014).

Since the Real Name Policy controversy, Facebook’s current name standards still rely on legal authority to verify the name, though they no longer use the language of ‘real name’. The policy states that the “name on your profile should be the name that your friends call you in everyday life” (*What names are allowed on Facebook?* 2019). However, it goes on to state “the name should also appear on an ID or a document on our ID list”. Why is it that Facebook needs an ID document to prove what name someone’s friends call them? Facebook couches an appeal to authority with the chattier preamble. Facebook wants to position themselves merely as the venue through which we talk to our friends, using the names our friends call us, but this is not the case. This example highlights a wider trend of responding to criticism with the appearance of change while the underlying policy remains the same. This is a discursual practice: to misdirect, to equivocate, to slightly change the language while maintaining the ideological function (Fairclough, 1995).



When discussing their rationale for the Sexual Solicitation policy change, Facebook correlates “sexual violence and exploitation” to “content [that] facilitates, encourages or coordinates sexual encounters between adults” (*Community standards*, 2018). On Facebook, one is allowed to discuss sexual violence and exploitation, but one is not allowed to discuss consensual sex. They “recognise the importance of Facebook as a place to...draw attention to sexual violence and exploitation” (*Community standards*, 2019). This distinction harkens back to Weitzer’s ‘moral crusade’ (2007). Sensationalised horror stories of sex trafficking victims can be used as currency to advance a political motive, but two adults arranging to meet and have consensual sex must not happen on Facebook. Facebook’s stance reflects the distorted public conversation around sex trafficking and sex work. Street-based sex workers and sex trafficking victims are a minority of all sex workers, but laws and organisations addressing them are disproportionately supported (Feldman, 2014). When only the stories of violence and exploitation are permitted, there is no room for sex workers to own their own narrative. “Awareness raisers can still count on a social hunger for lurid and detailed accounts, as well as a social order that restricts sex workers’ own opportunities to speak about the realities of their lives” (Grant, 2014, p. 37). Similar to how social media companies police women’s and queer people’s bodies and sexual agency—only tragedy is allowed. Our analysis must ask, whose voices are privileged? Whose are excluded? (Barker and Scheele, 2016). Sex workers cannot voice their lived experience but those with a vested interest in silencing and marginalising them have Facebook’s support to “create space for this conversation” (*Community standards*, 2019). Permitting only the voices of awareness-raisers sharing lurid traumas is rooted in misogyny and homophobia. Hegemony is maintained by controlling the narrative.

Additionally, as Sanders et al’s 2017 survey showed, the majority of those who engage in online sex work do so on a casual or part time basis. Yet the law wishes to criminalise part time sex workers on a full time basis, restricting their ability to engage with society even when they are not actively engaging in sex work, creating a criminalised class of person. “Laws against prostitution are used to target a class of people as whores whether or not they are selling sex, and in areas of their lives far outside what they do for a living” (Grant, 2014, p. 128).

The October 2018 Sexual Solicitation update was criticised heavily for its vague language and overly broad restrictions. Disallowing “sexualised slang” or “using sexual hints such as mentioning sexual roles... [or] sexual preference/sexual partner preference” basically censors any discussion of sex whatsoever. I am a member of a Facebook group called “Grindr

Aesthetics” where members post humorous topics on the trials and tribulations of gay dating, Grindr being a popular gay dating app. It is an important community for many members who express feeling like they do not have a similar outlet in their daily lives. Much of their content, and that of many similar groups, immediately became questionable under this policy. In December 2018, Facebook released an update to the Sexual Solicitation policy in response to the criticism. In it, Facebook places the onus on the media for simply misunderstanding the policy. “Recent media coverage and online conversation has revealed that our Sexual Solicitation policy is not entirely clear and has led to confusion and concern” (*Community standards*, 2018). They do not concede the possible validity of any criticism from those who did understand the policy but instead double down on its intent, to “ensure Facebook is not a place for coordinating sexual activity” (*Community standards*, 2018). Banning all sexual solicitation, whether for money or not, allows Facebook to shut the door on the issue and protect themselves from liability at the cost of fostering an open space where adults can discuss consensual sex. The December 2018 update also includes a call for critique, saying they “welcome feedback on how we can continue to improve and update our policies” (*Community standards*, 2018), yet there is not any kind of feedback submission form anywhere on the Community Standards webpage. Saying they welcome feedback without building any avenue to provide the feedback is merely a platitude to deflect criticism. Again, Facebook responds to criticism with lateral and cosmetic change.

Tumblr also changed its nudity and sexual content policies seemingly motivated in part by SESTA-FOSTA. As discussed in the introduction, they were also spurred on by a dispute with Apple. Apple deleted the Tumblr app from the App Store due their own liability concerns, showing regulation through law. This forced Tumblr to modify its behaviour and ban all adult content entirely. They apparently accepted their failure to moderate illegal CSE materials, and chose instead to eliminate all questionable content. Child sexual exploitation is a serious crime made in some ways easier by the Internet as the spread and share of CSE images can re-victimise the child. However, the vast majority of users will never see or interact with this media, so banning all legal adult content seems like an overreaction. Tumblr’s problem remains one of enforcement, not the law itself. They have exponentially increased the amount of content that must now be reviewed and deleted, when they were already struggling to keep up.

Tumblr’s new Adult Content policy bans all “images, videos, or GIFs that show real-life human genitals or female-presenting nipples—this includes content that is so photorealistic that it could

be mistaken for featuring real-life humans (nice try, though)” as well as “any content, including images, videos, GIFs, or illustrations, that depicts sex acts” (*Community guidelines*, 2018). The phrase ‘female-presenting nipples’ really stands out here. It is an awkward word choice, one that seems to not mean anything different from the standard ‘female nipples’ that is used in Instagram and Facebook’s policies. I can only postulate it is because Tumblr recognises their centrality to online queer communities and are attempting to mimic their gender-inclusive language, without much success. A person can be female-presenting, but one would be hard-pressed to define what makes a nipple female-presenting.

Tumblr is the only one of these policies that goes so far as to ban illustrations of sex. Possibly, this is because of Tumblr’s affordances for sharing and hosting of images, which positioned it as artists’ primary way of displaying their work online. “Tumblr users favoured communication through image...they relied less on text and more on the felt register of suggestive imagery, one of intimacy, assemblage, intensity, and aesthetic” (Cho, 2015, p. 44). This included a significant quantity of erotic fan art. Many of these artists have now decamped to Twitter, though Twitter lacks the same culture of the reblog—“the saturation and flow of images in the dashboard” (Cho, 2015, p. 54). Banning erotic illustration is an especially puritanical reaction to a problem of failed regulation.

Another regulation that arose in response to new law is that against nonconsensual pornography. Instagram, Twitter, Facebook, YouTube, and Tumblr all disallow nonconsensual sharing of nude or sexual images in firmly stated policies. They detail the different sorts of imagery this entails, from revenge porn to creepshots. Facebook filed their policy in the category of sexual exploitation and includes sharing or threatening to share private sexual conversations. YouTube’s policy includes media where other people are nonconsensually exposed to nudity, like “predatory exhibitionism” (*Policies and safety*, 2019). Since Twitter permits some sexual content and nudity, their policy indicates the ways in which they determine context. Outside of clearly nonconsensual intimate media, i.e. upskirt photos or nude images captioned with clear intent to harm the subject, Twitter requires the photo or video’s subject to report the image. These policies have become more serious over the last several years, presumably because awareness and public disapproval of the issue has become more widespread. Before the 2017 update, Tumblr’s nonconsensual pornography policy was written in a joking tone: “Don’t post content that violates anyone’s privacy, especially personally identifying or confidential information like...photos of your ex’s junk (no matter how remarkable)” (*Tumblr policy*, 2017).

After the update, it now explicitly bans nonconsensual pornography by name. As of July 2019, the UK and 46 US states have what are colloquially known as revenge porn laws, laws prohibiting nonconsensual pornography. These joking or colloquially-worded community guidelines were more common in the past before revenge porn was an issue with such widespread awareness. The social media nonconsensual pornography clauses are now in some cases more strongly-worded than the laws which they reflect. New York's revenge porn law includes language that the perpetrator must have "intent to cause harm" (Ellis, 2019). Twitter's policy bans all "explicit images or videos that were taken, appear to have been taken or that were shared without the consent of the people involved" (*Twitter rules and policies*, 2019). However, their instructions for who can report violations show how their priority is oriented towards a strong policy with enforcement a secondary concern. Only content that was obviously filmed without the subject's consent—"creepshots or upskirts"—or content containing text "that wishes/hopes for harm to come to those depicted" or containing personally identifying information can be reported by anyone. All other content must be reported by its subject "or an authorised representative, such as a lawyer" (*Twitter rules and policies*, 2019). This is yet another example of how a policy's wording can obfuscate its effect.

## 7.2 Norms

Norms regulate by encouraging socially-accepted behaviour and discouraging deviant behaviour. Social media users can construct their own sense of norms, both for behaving within small subcultures or counterpublics, and for expectations of the platform more generally. Or, the platform can instil norms in its users through encouraging interaction with their policies, often couched as 'community standards' or 'community guidelines'. Calling them community guidelines instead of rules is one way a social media company invites the user to follow them. The user is gently directed to follow the rules by all the ways the company makes them feel like good community citizens for compliance. People will even report each other for perceived violations, showing how deeply the norms can be instilled. "The ambiguity of the flag and its vocabulary is...an asset to the platform, leaving it room to honour flags in some cases and overrule them in others" (Gillespie, 2018, p. 102).

Facebook goes into much greater detail regarding specificities of allowed and disallowed content than any of the other social media platform's policies. Facebook's policy bans the vast majority of nude and sexual content, though in limited cases they have been pressured into

attempting to parse context and intent. Public pressure has changed their stance on nudity for educational and protest purposes, and female breasts in breastfeeding or mastectomy contexts (Gillespie, 2018). There are several bullet points in Facebook's policy that leads the reader to think there might have been a specific incident that resulted in that policy, such as: "Do not post images of...visible anus and/or fully nude close-ups of buttocks unless photoshopped on a public figure" (*Community standards*, 2019), or when, in June 2018, they added "snot" to their list of banned fetish content. Their exposed breast policy changes reflect an ongoing debate in society. A naked female breast is an effective form of protest precisely due to its shock value. Instagram uses similar language to Facebook in its no-nudity policy. "Photos of post-mastectomy scarring and women actively breastfeeding are allowed" (*Help centre*, 2019). Facebook allows nipples as protest, while Instagram does not. The female nipple is only permitted in a context of tragedy or motherhood. Again we must ask of this discourse, "What is included and excluded, privileged and silenced?" (Barker and Scheele, 2016, p. 201). What is the text doing here? It is making plain the hierarchy of value assigned to bodies. In order to challenge this double standard, transgender woman Courtney Demone started posting topless photos of herself on a regular basis to Facebook and Instagram. The challenge was to see at which her developing breasts became 'female' enough to be censored (Demone, 2015). This experiment highlighted the misogyny inherent in these no-nipple policies. Cisgender women's bodies cannot exist in public without being sexualised, and "people with differing bodies are sexualised, fetishized, exoticized and shamed" (Demone, 2015).

Both Facebook and Instagram reference a "community" or "diverse community" to enforce a sense of the imagined public. It is purposefully unclear which community is uncomfortable with nudity or sexual content. All of these companies like to describe their user base as a community, which is clearly intended to encourage the users to take on a sense of ownership of their experience on the platform. A user is not a lone individual interfacing with a website from their device, they are a member and participant in a group of people. People who feel they have a stake in the well-being and success of a community are more likely to manage their own behaviour and police the behaviour of others they see to be outliers. Instagram says, "We created the Community Guidelines so you can help us foster and protect this amazing community" (*Help centre*, 2019). The YouTube policy begins the same way: "When you use YouTube, you join a community of people from all over the world. Every cool, new community feature on YouTube involves a certain level of trust. Millions of users respect that trust and we trust you to be responsible too" (*Policies and safety*, 2019). YouTube invokes all the other

people who have no problem following their rules, in a way that sounds like ‘look, they’re all doing it, why can’t you?’ Saying ‘we trust you to be responsible’ makes that sense of ownership and call to action explicit.

Calling the policy documents ‘standards’ or ‘guidelines’ instead of ‘rules’ or ‘regulations’ is a way to give the rules a more positive spin. They are not truly mere guidelines, as they clearly list actions that are not allowed and will result in consequence, but naming them such gives the reader a sense of freedom. The company is gently shaping and guiding our online experience instead of directing it. When one of these documents does actually invoke the word ‘rules’ it establishes a different tone. “Our rules are to ensure all people can participate in the public conversation freely and safely” (*Twitter rules and policies*, 2019). Twitter’s approach is aligned with their stated mission to “serve the public conversation”. They take a slightly different approach from Facebook, Instagram, Tumblr, and YouTube due to Twitter’s brand positionality as a hub for news and political discussion.

### **7.3 Architecture**

Architecture regulates by creating the parameters of regulable space in the first place, and by deleting, hiding, or filtering that which falls outside those parameters. “Code is a regulator in cyberspace because it defines the terms upon which cyberspace is offered. And those who set those terms increasingly recognise the code as a means to achieving the behaviours that benefit them best” (Lessig, 2000, p. 84). Code, or architecture, regulates not by delivering consequences to the regulated, but by prohibiting the transgression before it can happen. Regulation through architecture is a reflection of the ‘technologization of discourse’ (Fairclough, 1995). This is a concept that builds on Foucault’s ‘technologies of power’. The technologization of discourse is a “top-down intervention to change discursive practices and restructure hegemonies within orders of discourse” (Fairclough, 1995, p. 87). Building a platform for speech and then regulating through code which speech is even visible is a distinctly top-down intervention.

The results of Facebook’s October 2018 Sexual Solicitation policy are unclear due to its recency, the impossibility of collecting data from Facebook, and the way Facebook reports violations to its users. Users are generally not told what policy their deleted content violated (West, 2018). Sex workers and adult performers often believe their online presences are unfairly

and persistently targeted for deletion, shadowbanning, etc., even when they attempt to stringently follow the stated policies (Mac and Smith, 2018). Shadowbanning is a complicated idea that is hard to explicitly evidence. Shadowbanning is when social media algorithms will de-index or de-prioritise the content of certain users, limiting their visibility and reach. Therefore the user is not banned, but they are existing outside of the platform at large. We do know that social media companies use algorithms to determine what content is put into the news feed or timeline and in what order. A scroll down the Facebook News Feed will show how their system values video, sponsored posts, and other high engagement content. “Communities, exchange, and conversation all flourish in a certain type of space; they are extinguished in a different type of space” (Lessig, 2000, p. 85). Twitter uses Safe Search to segregate ‘sensitive media’ and Tumblr also experimented with a Safe Mode, though it has since been scrapped in favour of prohibiting adult content entirely. Shadowbanning is problematic because it prohibits certain groups of people from participating in social media, even outside of their professional lives. These architectural decisions cast a wide net and do not differentiate between the profile of an adult performer, who operates legally, and other kinds of sex workers, whose purview is more generally illegal.

In 2013, Tumblr was bought by Yahoo, and then both were acquired by Verizon in 2017. These companies all integrated under the Oath brand umbrella. After Verizon’s purchase came Tumblr’s first attempt to regulate adult content by introducing ‘Safe Mode’, which supposedly allowed users to filter their Tumblr experience. Safe Mode was an opt-in feature for adults and automatically enabled on the accounts of minors, or at least all who had their age set to under 18 (Perez, 2017a). When Tumblr rolled out Safe Mode in 2017, they included all nudity in the ‘sensitive’ media category even when “in an artistic, educational, or photojournalistic context” (*Community Guidelines*, 2017). However, Safe Mode almost immediately came under fire for filtering innocuous LGBTQIA content. Tags like ‘gay’ or ‘trans’ were completely blocked. Tumblr stated that this was unintentional (Perez, 2017b). As established in the Sex and Sexuality Online chapter, Tumblr was a nexus for queer/LGBTQIA community, so even this supposedly unintentional censorship was a big blow to the user base. Users felt this was an example of LGBTQIA content being targeted as explicit no matter how benign it might be. Instances like this show how different communities with different norms can clash. These are the struggles inherent in power relations between the hegemonic norms and the norms of a counterpublic. Tumblr’s queer users who reacted to the Safe Mode gaffe are the networked counterpublic here. They organised around this issue and used networked technologies to protest a decision

they felt ran counter to their values. The debate even spilled out from Tumblr to other social media platforms as users shared examples of the kinds of content that were getting erroneously filtered. The use of filtering for some users places this example fairly clearly under regulation by architecture, but the timing of the filtering is so convenient that it seems like regulation through the market is in play as well.

Tumblr's initial attempts to regulate adult content in the wake of the December 2018 policy change were unsuccessful, leading to some nearly humorous algorithmic mistakes. The post written by the Tumblr Staff account, detailing the policy change, was itself flagged under the new policy for explicit material. Speculation was rampant about the function of the new nudity-detection algorithm. While these programs are decent at detecting partial nudity, there is little way to differentiate between a shirtless man (allowed) and a shirtless woman (not allowed). To a program, the swath of skin-toned pixels looks the same. For the first several weeks after implementation, it was possible to avoid flagging by applying a green filter to an image of nudity. There was speculation that the policy had been implemented without first using a corpus of training data, and that the machine learning was happening live. Tumblr users took to other social media platforms to show the most absurd examples of flagged posts, including a photo of a hand, a photo of two fully-dressed people, a photo of a mannequin, a photo of a chocolate squirrel, and a patent application illustration of a tire (Burstein, 2018).

Twitter allows adult nudity, though they have a Safe Search feature intended to limit its reach. Safe Search is part of Twitter's overall stated mission of prioritising "high-quality tweets". Safe Search works by filtering out media it deems explicit, irrelevantly tagged, or spammy from search results. De-indexing spammy or abusive accounts is generally noncontroversial, as most people do not want their Twitter experience inundated with spam or threats. However, this burying of accounts that create and share adult content leads to user perception of bias. The content is allowed under Twitter's rules, but it is not granted equal access to the platform. When Twitter decides to deindex the accounts of pornographers, sex workers, cam girls, and the like, that is a form of regulation. Twitter using its search algorithms to limit their reach is a perfect example of the code is law maxim. Prohibiting them from expanding their Twitter audience through the architectural decisions in the search algorithm also punishes adult performers by preventing them from growing their businesses or harming their livelihood. Twitter's 'sensitive media' policy places adult content in the same category as "graphic violence", "gratuitous gore", and "hateful imagery" (*Twitter rules and policies*, 2019). Users are required to mark their own



content as sensitive, placing it behind an interstitial barrier. A potential viewer of sensitive media must first click an acknowledgement message before the media is displayed.

## 7.4 Markets

Markets regulate by monetarily incentivizing certain decisions and disincentivizing others. “Advertisers reward popular sites; online services drop low-population forums. These behaviours are all a function of market constraints and market opportunity” (Lessig, 2000, p. 124). Markets also regulate because people have to make money to survive under capitalism. People will be discouraged from behaviours that affect their earning potential, or conversely, people are forced into behaviours they would not exhibit outside of dire economic circumstances. SESTA-FOSTA regulates through the market as well as by law. “Through such demands [against Craigslist, Backpage, and other venues for sex workers’ ads], reformers take away sex workers’ power to make these decisions about their own labour...losing ad venues means losing control of how they negotiate at work” (Grant, 2014, p. 65). Restricting sex workers’ access to safer labour opportunities is regulation through the market. Legislators seek to control women’s and queer people’s bodies and labour. By banning something like ‘implicit sexual solicitation’, Facebook is complicit in this regulation by market. “Discoursal practice is a facet of the struggle which contributes in varying degrees to the reproduction or transformation of the existing order of discourse, and through that of existing social and power relations” (Fairclough, 1995, p. 77). The text of ‘implicit sexual solicitation’ becomes discoursal practice because of its function to maintain these unequal power relations.

YouTube allows content creators to monetise their videos by joining the YouTube Partner Program and hosting ads on their videos through Google AdSense. The YouTube Partner Program is not available to all creators, however. YouTube has strict ‘advertiser-friendly content guidelines’ that delineate which kinds of videos can be monetised. Among the topics that are not suitable for advertisers are: “inappropriate language, violence, adult content, harmful or dangerous acts, hateful content, incendiary and demeaning, recreational drugs and drug-related content, tobacco-related content, firearms-related content, controversial issues and sensitive events, adult themes in family content” (*Policies and safety*, 2019). These policies create two tiers of YouTube creators. The market regulates by only permitting non objectionable, clean videos to generate revenue for their creators. Considering YouTube is the only game in town, this is effectively censorship. Even bleeped-out profanity is not allowed. All “sexually suggestive

content” is not available for monetisation (*Policies and safety*, 2019). This includes “content related to the sex industry” and “content discussing intimate sexual experiences” (*Policies and safety*, 2019). That means most of the communities I have discussed here, whether queer young adults on Tumblr or sex worker activism blogs, would not be allowed access to this tier of YouTube membership. This is regulation through the market. While technically some of these non-monetizable topics are permitted on YouTube as a whole, creators are strongly disincentivized from their production due to their ineligibility under the Partner Program. Videos are judged for ‘advertising-friendliness’ with an automated system. The machine looks at the title, description, thumbnail image, and tags, as well as using voice and image recognition on the content of the video. This regulation through architecture. Once a video has been deemed unsuitable by the system, the creator can request a human moderator review the decision. However, YouTube does not offer all its creators the right to human intervention equally. YouTube creators with more than 10,000 subscribers get automatic access to the human moderation team, jumping the queue. They also can have the videos reviewed before they are uploaded to the public, so they do not lose that crucial first week revenue, when the largest number of views are accumulated. Creators with less than 10,000 subscribers can only get their ‘advertiser-friendliness’ reviewed after uploading and publishing the video. Even if the decision is reversed, that creator will still have lost the majority of the revenue from that video. In these ways, YouTube regulates the output of its creators by controlling their ability to make money. Hegemony is a process that works first in the personal and societal realms (Fairclough, 1995). The choices of who can be monetised to speak, and about what, are a discursive practice that maintains hegemony.

## 8. Conclusion

Social media regulation affects our online interactions, which informs and is informed by the offline world. It is a reflection of what our society values. Discourses are shaped by what speech is allowed. Hegemony is maintained by disseminating ideology through discourse. Hegemony is challenged through the formation of counterpublics who use their internal discourse to build solidarity and gain senses of collective identity. These conflicting discourses then shift the balance of power relations and produce tension that fosters resistance to hegemony. Social media policies delineate permissible discourse, and therefore have an outsized effect on these societal processes.

## **8.1 How do social media companies regulate adult content?**

Social media companies regulate adult content by invoking the law, fostering norms, writing code, and through the influence of the market. Over time, these processes have become more complex, more opaque, and in many cases more restrictive. In some cases, public pressure has resulted in policy change, but usually not in a meaningful way. When Facebook faced criticism for their Real Name Policy, they basically just expanded the forms of ID they accept. Shifting cultural attitudes are reflected in social media policies, as are political changes. New 'revenge porn' laws have arisen concurrently with social media's revenge porn policies. SESTA-FOSTA also dramatically altered the online landscape. When enforcing their policies, social networks use a mix of algorithmic content detection and human moderation. Neither is a perfect system, and even when working together, regulation is inconsistently applied and sporadically enforced, leading to a lot of frustration for the user.

## **8.2 Can we see each mode of Lessig's regulation working on and through social media policies?**

Lessig's modalities of regulation mapped well on to the case study of social media policies. When Lessig first published *Code is Law* in 2000, the landscape of the Internet was very different than it was today. However, despite the emergence of technologies he could not have predicted, his frameworks of regulation are still evident. Policies are reflections of one set of norms that they then impose on everyone else. Norms and architecture are the most evident modalities in the case of social media, but laws and markets have their influence as well.

## **8.3 What effect does regulating adult content have on marginalized people?**

When considering regulation, censorship, and free speech in the virtual world, adult content may seem like a lesser worry. However, "sexual politics...are fundamental technologies of power that secure the reproduction of capitalism. They do so in a subtle manner by deploying heteronormative and racialized phantasms of 'sexual normality' and by inciting a desire for organizing consensus to the capital mode of production and the state" (Ludwig, 2018, p. 137). Heteronormativity and capitalism reinforce and reproduce each other. Heteronormativity is a pillar of capitalism due to how capitalism must enforce a hierarchy of power in order to

function; capitalism is a pillar of heteronormativity because it has transformed our relational and family structures along the same lines of labour and capital. We can see this reflected in what expressions of sexuality are censored by social media regulations. When regulation is inflicted upon a space, the most vulnerable members of the space are the first and most severely affected.

Hopefully, what this dissertation has not done is divide my audience on the issue of sex work. However you feel about sex work, we can agree that reducing the harm and violence that sex workers experience is good. In these cases of Internet regulation I have presented, the good intentions of law-makers or coders has led to more harm, not less. Under capitalism, all labour is exploitation, so criminalising sex work as some further degree of exploitation is ineffective. Criminalisation leads to greater danger of exploitation, as workers are left without support or resources for help. Of course, regulating the labour of sex workers has never been about protecting them from exploitation, or the regulators would base their policies off evidence that decriminalisation is the best way to reduce harm.

## **9. Reflections and Recommendations**

Future studies could use the same framework to examine regulation of different sorts of content. While my focus was adult content, another researcher could analyse regulation on hate speech, far right groups, anti-vaxxers, propagandists, fake news propagators, and so on. There has been immense change in all of these realms over the last several years as well. Clearly there are specific affordances of social media that allow widespread disinformation and propaganda campaigns—how and should these be regulated? What are these affordances, from Lessig's 'code is law' perspective? There is an alternate perspective to examine here: when does regulation work, and what does that look like? Lessig and I agree, a web space without regulation is impossible, both because it would be unusably inundated with spam and because the architectural decisions necessary to build the space impose some regulation. Additionally, Lessig's modes of regulation can be used as a framework for other investigations into online communities outside of social media.

Any investigation into social media is hampered by the platforms' opacity. Any conclusions we draw can only be based on what the user sees in the front end experience of the platform; companies are committed to making that experience inexplicable. No two users see the same

home page or are recommended to or advertised to in the same way. Even professional users of social media can only use educated guesses to try and cater to the algorithms. Only social media companies can know if something like shadowbanning is as widespread a practice as is believed by its supposed victims.

Unfortunately, the state of the Internet today might be the problem. We all exist in cyberspace mostly within these corporate silos. Social media is not regulated because it is ethically imperative to do so, it is regulated because social media must maintain its advertiser-friendliness. It is essentially impossible to craft one set of rules that can be applied regardless of culture, local law, and context. Facebook's Community Standards must work for all of their several billion users. So we end up with the policies we have examined here today: uselessly vague in some regards, overly restrictive in other areas, unequally applied and literally traumatic to enforce for underpaid, outsourced content moderators. Perhaps the only solution is a more fragmented online social experience—something we are already seeing with the rise of group chat programs like Discord. Dismantling the hegemonic control of these five or so companies that control most of our online experience is worth considering. And remember, folks, sex work is work and all work is bad.

## References

- Angwin, J. (2017) 'Facebook's secret censorship rules protect white men from hate speech but not black children', *ProPublica*, 28 June. Available at: <https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms> (Accessed: 4 July 2019).
- Axelrod-Contrada, J. (2007) *Reno v. ACLU: Internet censorship*. Tarrytown, NY: Marshall Cavendish Benchmark.
- Barako, T. E. 'Finding a DMCA safe harbor in the turbulent sea of online copyright liability: *A&M Records, Inc. v Napster, Inc.*'. *Jurimetrics* 42(1), pp. 1-20.
- Barker, M. and Scheele, J. (2016) *Queer: A graphic history*. London: Icon Books.
- boyd, d. (2010) 'Social network sites as networked publics: Affordances, dynamics, and implications', in Papacharissi, Z. (ed.) *A networked self: Identity, community, and culture on social network sites*. Abingdon, Oxon: Routledge, pp. 39-57.
- Brandon, R. (2014) 'Facebook's Report Abuse button has become a tool of global oppression', *The Verge*, 2 September. Available at: <https://www.theverge.com/2014/9/2/6083647/>

- facebook-s-report-abuse-button-has-become-a-tool-of-global-oppression (Accessed: 4 July 2019).
- Bruns, A. and Moe, H. (2014) 'Structural layers of communication on Twitter', in Weller, K. et al. (eds.) *Twitter and society*. New York: Peter Lang, pp. 15-28.
- Burger, W. E. and the Supreme Court of the United States (1972) *U.S. Reports: Miller v. California*, 413 U.S. 15. Available at: <https://www.loc.gov/item/usrep413015/> (Accessed: 1 August 2019).
- Burstein, S. (2018) 4 December. Available at: [https://twitter.com/design\\_law/status/1069967727140528128](https://twitter.com/design_law/status/1069967727140528128) (Accessed: 25 July 2019).
- Caplan, R. (2018) *Content or context moderation? Artisanal, community-reliant, and industrial approaches*. New York: Data and Society. Available at: [https://datasociety.net/wp-content/uploads/2018/11/DS\\_Content\\_or\\_Context\\_Moderation.pdf](https://datasociety.net/wp-content/uploads/2018/11/DS_Content_or_Context_Moderation.pdf) (Accessed: 4 July 2019).
- Cavalcante, A. (2018) 'Tumbling into queer utopias and vortexes: Experiences of LGBTQ social media users on Tumblr'. *Journal of Homosexuality* pp. 1-21.
- Cockbain, E. and Bowers, K. (2019) 'Human trafficking for sex, labour, and domestic servitude: How do key trafficking types compare and what are their predictors?' *Crime, Law and Social Change*, pp. 1-26.
- Charlesworth, J. C. (2011) 'The moral of the story: What *Grokster* has to teach about the DMCA'. *Stanford Technology Law Review* 6, pp. 1-14.
- Cho, A. (2015) 'Queer reverb: Tumblr, affect, time', in Hillis, K., Paasonen, S. and Petit, M. (eds.) *Networked affect*. Cambridge, MA: MIT Press, pp. 43-57.
- Community guidelines* (2018). Available at: <https://www.tumblr.com/policy/en/community> (Accessed: 30 July 2019).
- Community standards* (2018) Available at: <https://www.facebook.com/communitystandards/recentupdates/> (Accessed: 26 July 2019).
- Community standards* (2019) Available at: <https://www.facebook.com/communitystandards/> (Accessed: 2 March 2019).
- Cooper, D. (2018) 'FOSTA-SESTA's real aim is to silence sex workers online', *EnGadget*, 11 April. Available at: <https://www.engadget.com/2018/04/11/fosta-sesta-silencing-sex-workers/> (Accessed: 4 March 2019).
- DecrimNY (2019) 26 July. Available at: <https://twitter.com/DecrimNY/> (Accessed: 26 July 2019).
- Demone, C. (2015) 'Do I have boobs now?', *Mashable*, 30 September. Available at: <https://mashable.com/2015/09/30/do-i-have-boobs-now/> (Accessed: 27 July 2019).

- Dorman, S. (2019) 25 July. Available at: <https://twitter.com/SarahxDorman/status/1154462415293038597> (Accessed: 26 July 2019).
- Ellis, E. G. (2019) 'New York's revenge porn law is a flawed step forward', *Wired*, 24 July. Available at: <https://www.wired.com/story/new-york-revenge-porn-law/> (Accessed: 27 July 2019).
- Fairclough, N. (1995) *Critical discourse analysis: The critical study of language*. New York; London: Longman Publishing.
- Fairclough, N. (2001) in Wodak, R. and Meyer, M. (eds) *Methods of critical discourse analysis*. London, UK: Sage Publications, pp. 121-138.
- Fairclough, N. (2015) *Language and power*. 3rd edn. Abingdon, Oxon: Routledge.
- Feldman, V. (2014) 'Sex work politics and the Internet: Carving out political space in the blogosphere', in Showden, C. R. and Majic, S. (eds) *Negotiating sex work: Unintended consequences of policy and activism*. Minneapolis, MN; London: University of Minnesota Press, pp. 243-266.
- Fraser, N. (1990) 'Rethinking the public sphere: A contribution to the critique of actually existing democracy'. *Social Text* 25, pp. 56-80.
- Gerrard, Y. (2018) 'Beyond the hashtag: Circumventing content moderation on social media'. *New Media and Society* 20(12), pp. 4492-4511.
- Gillespie, T. (2018) *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. New Haven, CT: Yale University Press.
- Godwin, M. (2003) *Cyber rights: Defending free speech in the digital age*. Cambridge, MA: MIT Press.
- Grant, M. G. (2014) *Playing the whore: The work of sex work*. London; New York: Verso Books.
- Habermas, J. (1991) *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. Burger, T. and Lawrence, F. (trans.) Cambridge, MA: MIT Press.
- Heineman, J. and Wagner, B. (2018) 'The sex trafficking panic is based on myths', *Buzzfeed News*, 18 April. Available at: <https://www.buzzfeednews.com/article/jennyheineman/sex-trafficking-myths-sesta-fosta> (Accessed: 11 June 2019).
- Help centre (2019) Available at: <https://help.instagram.com/477434105621119> (Accessed: 26 July 2019).
- Hintz, A. (2016) 'Restricting digital sites of dissent: Commercial social media and free expression'. *Critical Discourse Studies* 13(3), pp. 325-340.
- Home Office (2019) *Country policy and information note - Nigeria: Trafficking of women*.

- London: UK Visas and Immigration. Available at: <https://www.gov.uk/government/publications/nigeria-country-policy-and-information-notes> (Accessed 28 June 2019).
- Jenkins, H., Ito, M. and boyd, d. (2016) *Participatory culture in a networked era: A conversation on youth, learning, commerce, and politics*. Cambridge, UK: Polity Press.
- Jeong, S. (2017) 'Sex trafficking bill is turning into a proxy war over Google', *The Verge*, 14 September. Available at: <https://www.theverge.com/2017/9/14/16308066/sex-trafficking-bill-sesta-google-cda-230> (Accessed: 11 June 2019).
- Johnson, T. (2017) '21st Century Fox backs sex trafficking bill opposed by major Internet firms', *Variety*, 13 September. Available at: <https://variety.com/2017/politics/news/sex-trafficking-act-21st-century-fox-1202557318/> (Accessed: 11 June 2019).
- Jones, A. (2015) 'Sex work in a digital era'. *Sociology Compass* 9(7), pp. 558-570.
- Kaplan, A. M. and Haenlein, M. (2010) 'Users of the world, unite! The challenges and opportunities of Social Media'. *Business Horizons* 53(1), pp. 59-68.
- Kinney, E. (2014) 'Raids, rescues, and resistance: Women's rights and Thailand's response to human trafficking', in Showden, C. R. and Majic, S. (eds) *Negotiating sex work: Unintended consequences of policy and activism*. Minneapolis, MN; London: University of Minnesota Press, pp. 145-170.
- Kucan, L. (2011) 'Approximating the practice of writing the dissertation literature review'. *Literacy Research and Instruction* 50(3), pp. 229-240.
- Kuzma, A. (2013) 'A letter to Congress: The Communications Decency Act promotes human trafficking'. *Children's Legal Rights Journal* 34(1), pp. 23-58.
- Langlois, G. et al (2009) 'Networked publics: The double articulation of code and politics on Facebook'. *Canadian Journal of Communication* 34, pp. 415-434.
- Leary, M. G. (2018) 'The indecency and injustice of section 230 of the Communications Decency Act'. *Harvard Journal of Law and Public Policy* 41(2), pp. 553-622.
- Lessig, L. (2000) *Code and other laws of cyberspace*. New York: Basic Books.
- Lil Miss Hot Mess (2015) 'Selfies and side-eye: Drag queens take on Facebook'. *Studies in Gender and Sexuality* 16, pp. 144-146.
- Ludwig, G. (2018) 'Queering capitalism'. *Krisis Journal for Contemporary Philosophy* 2, pp. 135-138.
- Mac, J. and Smith, M. (2018) *Revolting prostitutes: The fight for sex workers' rights*. London; New York: Verso Books.
- McKinney, C. J. (2019) 'Women trafficked into prostitution are "wealthy" and don't need asylum, Home Office claims'. *Free Movement*, 28 June. Available at: <https://www.freemovement.org>



- org.uk/women-trafficked-into-prostitution-are-wealthy-and-dont-need-asylum-home-office-claims/ (Accessed: 28 June 2019).
- Miller-Young, M. (2010) 'Putting hypersexuality to work: Black women and illicit eroticism in pornography'. *Sexualities* 13(2), pp. 219-235.
- Mondin, A. (2017) 'Tumblr mostly, great empowering images: Blogging, reblogging, and scrolling feminist, queer, and BDSM desires'. *Journal of Gender Studies* 26(3), pp. 282-292.
- Muise, A. (2011) 'Women's sex blogs: Challenging dominant discourses of heterosexual desire'. *Feminism and Psychology* 21(3), pp. 411-419.
- Neumayer, C. and Valtysson, B. (2013) 'Tweet against Nazis? Twitter, power, and networked publics in anti-fascist protests'. *MedieKultur* 55, pp. 3-20.
- Newton, C. (2019a) 'The trauma floor: The secret lives of Facebook moderators in America', *The Verge*, 25 February. Available at: <https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona> (Accessed: 2 March 2019).
- Newton, C. (2019b) 'Bodies in seats', *The Verge*, 19 June. Available at: <https://www.theverge.com/2019/6/19/18681845/facebook-moderator-interviews-video-trauma-ptsd-cognizant-tampa> (Accessed: 29 June 2019).
- Oakley, A. (2016) 'Disturbing hegemonic discourse: Nonbinary gender and sexual orientation labeling on Tumblr'. *Social Media and Society* 3, pp. 1-12.
- O'Connell, J. (2019) 'Facebook's dirty work in Ireland: "I had to watch footage of a person being beaten to death"', *The Irish Times*, 30 March. Available at: <https://www.irishtimes.com/culture/tv-radio-web/facebook-s-dirty-work-in-ireland-i-had-to-watch-footage-of-a-person-being-beaten-to-death-1.3841743> (Accessed: 28 June 2019).
- Perez, S. (2017a) 'Tumblr rolls out new content filtering tools with the launch of 'Safe Mode'', *TechCrunch*, 20 June. Available at: <https://techcrunch.com/2017/06/20/tumblr-rolls-out-new-content-filtering-tools-with-launch-of-safe-mode/> (Accessed: 4 March 2019).
- Perez, S. (2017b) 'Tumblr says it fixed the 'Safe Mode' glitch that hid innocent posts, including LGBTQ+ content', *TechCrunch*, 24 June. Available at: <https://techcrunch.com/2017/06/24/tumblr-says-it-fixed-the-safe-mode-glitch-that-hid-innocent-posts-including-lgbtq-content/> (Accessed: 22 July 2019).
- Pizzi, M. (2014) 'The Syrian opposition is disappearing from Facebook', *The Atlantic*, 4 February. Available at: <https://www.theatlantic.com/international/archive/>

- 2014/02/the-syrian-opposition-is-disappearing-from-facebook/283562/ (Accessed: 4 July 2019).
- Policies and safety* (2019) Available at: <https://www.youtube.com/intl/en-GB/yt/about/policies/#community-guidelines> (Accessed 26 July 2019).
- Porter, J. (2018) 'Tumblr was removed from Apple's App Store over child pornography issues', *The Verge*, 20 November. Available at: <https://www.theverge.com/2018/11/20/18104366/Tumblr-ios-app-child-pornography-removed-from-app-store> (Accessed: 4 March 2019).
- Pratt, M. L. (1991) 'Arts of the contact zone'. *Profession* 91, pp. 33-40.
- Pybus, J. (2015) 'Accumulating affect: Social networks and their archives of feelings', in Hillis, K., Paasonen, S. and Petit, M. (eds.) *Networked affect*. Cambridge, MA: MIT Press, pp. 235-249.
- Renegade, R. and Pottenger, K. (2019) 'Sex work is work'. *New Labor Forum* 28(1), pp. 98-102.
- Renninger, B. J. (2015) "'Where I can be myself...where I can speak my mind": Networked counterpublics in a polymedia environment'. *New Media and Society* 17(9), pp. 1513-1529.
- Restricted content* (2019) Available at: <https://play.google.com/about/restricted-content/> (Accessed: 4 March 2019).
- Roodenburg, J. M. (2001) "'Son of CDA": The constitutionality of the Child Online Protection Act of 1998'. *Communication Law & Policy* 6(1), pp. 227-257.
- Sanders, T. et al (2017) *Beyond the gaze: Summary briefing on Internet sex work*. Leicester, UK: University of Leicester Department of Criminology. Available at: <https://www.beyond-the-gaze.com/briefings/> (Accessed: 6 May 2019).
- Sanders, T. et al (2018) *Internet sex work: Beyond the gaze*. Cham, Switzerland: Palgrave Macmillan.
- Schmidt, J. (2014) 'Twitter and the rise of personal publics', in Weller, K. et. al. (eds.) *Twitter and society*. New York: Peter Lang, pp. 3-14.
- Smith, B. H. (2000) 'To filter or not to filter: The role of the public library in determining internet access'. *Communication Law and Policy* 5(3), pp.385-422.
- Steimle, S. (2019) 'New laws forced sex workers back on SF streets, caused 170% spike in human trafficking', *KPIX 5*, 3 February. Available at: <https://sanfrancisco.cbslocal.com/2019/02/03/new-laws-forced-sex-workers-back-on-sf-streets-caused-170-spike-in-human-trafficking/> (Accessed: 9 May 2019).

- Tiidenberg, K. (2013) 'How does online experience form our sense of self? NSFW blogger identity narratives', in Allaste, A. A. (ed.) *Back in the west: Changing lifestyles in transforming societies*. New York: Peter Lang, pp. 177-202.
- Tufekci, Z. (2017) *Twitter and tear gas: The power and fragility of networked protest*. New Haven, CT; London: Yale University Press.
- Tumblr policy* (2017) Available at: <https://github.com/tumblr/policy> (Accessed: 30 July 2019).
- Twitter rules and policies* (2019) Available at: <https://help.twitter.com/en/rules-and-policies> (Accessed: 26 July 2019).
- Van Dijck, J. (2013) *The culture of connectivity: A critical history of social media*. Oxford; New York: Oxford University Press.
- Vengattil, M. and Dave, P. (2019) 'Some Facebook content reviewers in India complain of low pay, high pressure', *Reuters*, 28 February. Available at: <https://in.reuters.com/article/facebook-content-india/some-facebook-content-reviewers-in-india-complain-of-low-pay-high-pressure-idINKCN1QH15E> (Accessed: 2 March 2019).
- Weitzer, R. (2007) 'The social construction of sex trafficking: Ideology and institutionalization of a moral crusade'. *Politics and Society* 35(3), pp. 447-475.
- West, S. M. (2018) 'Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms'. *New Media and Society* 20(11), pp. 4366-4383.
- Wodak, R. and Meyer, M. (eds) (2001) *Methods of critical discourse analysis*. London, UK: Sage Publications.
- Wood, E. A. (2008) 'Consciousness-raising 2.0: Sex blogging and the creation of a feminist sex commons'. *Feminism and Psychology* 18(4), pp. 480-487.

# Appendix: Social Media Policies

## Facebook 2018

### 15. Sexual Solicitation

December 2018

#### III: Objectionable Content

Do not:

Attempted coordination of or recruitment for adult sexual activities including but not limited to:

Filmed sexual activities  
Pornographic activities, strip club shows, live sex performances, erotic dances  
Sexual, erotic, or tantric massages

Explicit sexual solicitation which we define as offering or requesting things, including but not limited to the following:

Sex or sexual partners  
Sex chat or conversations  
Nude images

Implicit sexual solicitation, which we define as an offer or request to engage paired with at least one of the following elements, none of which violate our Community Standards on their own:

Vague suggestive statements, such as “looking for a good time tonight”  
Sexualized slang  
Using sexual hints such as mentioning sexual roles, sex positions, fetish scenarios, sexual preference/sexual partner preference, state of arousal, act of sexual intercourse or activity (sexual penetration or self-pleasuring), commonly sexualized areas of the body such as the breasts, groin, or buttocks, state of hygiene of genitalia or buttocks

Content (hand drawn, digital, or real-world art) that may depict explicit sexual activity or suggestively posed person(s)

An offer or ask for other adult activities such as:

Commercial pornography  
Partners who share fetish or sexual interests

Sexually explicit language that adds details and goes beyond mere naming or mentioning of:

A state of sexual arousal (wetness or erection)  
An act of sexual intercourse (sexual penetration, self-pleasuring or exercising fetish scenarios)

*Recent media coverage and online conversation has revealed that our Sexual Solicitation policy is not entirely clear and has led to confusion and concern.*

*To that end, we have revised the language to clarify our policy. Under the policy, implicit sexual solicitation is not allowed. We define implicit sexual solicitation as an offer or ask to engage paired with at least one suggestive element, which would not violate on its own.*

*Deciding where to draw the line is always a challenge and we welcome feedback on how we can continue to improve and update our policies. Our approach has been to ensure Facebook is not a place for coordinating sexual activity.*

## **CURRENT VERSION**

**October 2018**

### **III: Objectionable Content**

#### **Policy Rationale**

As noted in Section 8 of our Community Standards (Sexual Exploitation of Adults), people use Facebook to discuss and draw attention to sexual violence and exploitation. We recognize the importance of and want to allow for this discussion. We draw the line, however, when content facilitates, encourages or coordinates sexual encounters between adults. We also restrict sexually explicit language that may lead to solicitation because some audiences within our global community may be sensitive to this type of content and it may impede the ability for people to connect with their friends and the broader community.

Do not post:

Content that attempts to coordinate or recruit for adult sexual activities including but not limited to:

Filmed sexual activities

Pornographic activities, strip club shows, live sex performances, erotic dances

Sexual, erotic, or tantric massages

Content that engages in explicit sexual solicitation by, including but not limited to the following, offering or asking for:

Sex or sexual partners

Sex chat or conversations

Nude images

Content that engages in implicit sexual solicitation, which can be identified by mentioning a sexual act and other suggestive elements such as any of the following:

Vague suggestive statements, such as “looking for a good time tonight”

Sexualized slang

Using sexual hints such as mentioning sexual roles, sex positions, fetish scenarios, sexual preference/sexual partner preference, state of arousal, act of sexual intercourse or activity (sexual penetration or self-pleasuring), commonly sexualized areas of the body such as the breasts, groin, or buttocks, state of hygiene of genitalia or buttocks

Content (hand drawn, digital, or real-world art) that may depict explicit sexual activity or suggestively posed person(s)

Content that offers or asks for other adult activities such as:

Commercial pornography

Partners who share fetish or sexual interests

Sexually explicit language that adds details and goes beyond mere naming or mentioning of:

A state of sexual arousal (wetness or erection)

An act of sexual intercourse (sexual penetration, self-pleasuring or exercising fetish scenarios)

*This policy, which disallows sexual solicitation, has long been in place under the headings of “Sexual Exploitation of Adults” and “Adult Nudity and Sexual Activity” (Section 14). We know sexual solicitation isn’t necessarily exploitative, and that the language we use and their organization under our Community Standards matter a lot. We also heard from our review teams that the policy as previously written wasn’t*

clear, so we sought to distinguish the solicitation policy with its own heading (See [Section 15: Sexual Solicitation](#)) and removed this language from the Sexual Exploitation of Adults section of our Community Standards.

This policy change went through our formal policy process in 2018, which included getting external input from a number of stakeholders.

The policy covers 1) coordination and recruitment of sexual activities, 2) explicit > sexual solicitation, and 3) implicit sexual solicitation, 4) offers or asks for pornography or sexual partners, and 5) sexually explicit language.

## Facebook 2019

### 8. Sexual Exploitation of Adults

#### Policy Rationale

We recognize the importance of Facebook as a place to discuss and draw attention to sexual violence and exploitation. We believe this is an important part of building common understanding and community. In an effort to create space for this conversation while promoting a safe environment, we remove content that depicts, threatens or promotes sexual violence, sexual assault, or sexual exploitation, while also allowing space for victims to share their experiences. We remove content that displays, advocates for, or coordinates sexual acts with non-consenting parties or commercial sexual services, such as prostitution and escort services. We do this to avoid facilitating transactions that may involve trafficking, coercion, and non-consensual sexual acts.

To protect victims and survivors, we also remove images that depict incidents of sexual violence and intimate images shared without permission from the people pictured. For additional information on these efforts, please visit [Using Technology to Protect Intimate Images and Help Build a Safe Community](#), as well as our [guide to reporting and removing intimate images shared without your consent](#).

#### Do not post:

In instances where content consists of any form of non-consensual sexual touching, crushing, necrophilia or bestiality, including:

- Depictions (including real photos/videos), or
- Advocacy (including aspirational and conditional statements), or
- Statements of intent, or
- Calls for action, or
- Participation by yourself or others to engage in any form of the above mentioned sexual acts.

Content that attempts to exploit people by any of the following:

- Coercing money, favors, or images from people by threats of exposure of their naked or semi-naked photos/videos  
Sharing imagery that fulfills all three of the following conditions:
  - Image is non-commercial or produced in a private setting
  - Person in the image is (near) nude, engaged in sexual activity, or in a sexual pose
  - Lack of consent to share the image is indicated by
- Vengeful context (for example, caption, comments, or page title)  
Independent sources (for example, media coverage or law enforcement record)  
A visible match between the person depicted in the image and the person who has reported the content to us  
The person who reported the content to us shares the same name as the person depicted in the image  
Sharing imagery of people or a person focusing on sexualized areas of the body such as the breasts, groin, or buttocks (also known as creepshots or upskirts) or focusing on people engaged in sexual activity. The following elements need to be present:
  - the focal point is on a sexualized area of the body or sexual activity, and the person in the image is clearly unaware
  - Threatening or stating an intent to share intimate imagery without consent
  - Soliciting intimate imagery to view or share without consent
  - Threatening or stating an intent to share private sexual conversations

Attempting to coordinate adult commercial sexual services or prostitution activities, such as requesting or offering or asking for rates for escort services and paid sexual fetish or domination services.

For the following content, we include a warning screen so that people are aware the content may be disturbing:

Narratives and statements that contain a depiction of non-consensual sexual touching (written or verbal) that includes details beyond mere naming or mentioning the act if:

Shared by the victim, or  
Shared by a third party (other than the victim) in support of victim or condemnation of act or for general awareness to be determined by context/caption

### **7. Child Nudity and Sexual Exploitation of Children**

#### Policy Rationale

We do not allow content that sexually exploits or endangers children. When we become aware of apparent child exploitation, we report it to the National Center for Missing and Exploited Children (NCMEC), in compliance with applicable law. We know that sometimes people share nude images of their own children with good intentions; however, we generally remove these images because of the potential for abuse by others and to help avoid the possibility of other people reusing or misappropriating the images.

We also work with external experts, including the [Facebook Safety Advisory Board](#), to discuss and improve our policies and enforcement around online safety issues, especially with regard to children.

Do not post:



Content that depicts participation in or advocates for the sexual exploitation of children, including (but not limited to)

Engaging in any sexual activity involving minors

Adults soliciting minors

Minors soliciting minors

Minors soliciting adults

Using our products and site functionality with the intention of sexualizing minors

Content that constitutes or facilitates inappropriate interactions with children, such as

Initiating unsolicited contact with minors (for example, private messages between stranger adults and minors)

Soliciting, displaying, sharing, or viewing imagery of nude, sexualized, or sexual activity with minors

Arranging real-world sexual encounters or obtaining sexual material from a minor directly

Displaying nudity to minors

Content (including photos, videos, real-world art, digital content, and text) that depicts

- Any sexual activity involving minors
- Minors in a sexual fetish context
- Minors with sexual elements, including (but not limited to):

Restraints

Focus on genitals

Presence of aroused adult

Presence of sex toys

Sexualized costume

Stripping

Staged environment (for example, on a bed) or professionally shot (quality/focus/angles)

Open-mouth kissing with minor or adult

Content (including photos, videos, real-world art, digital content, and verbal depictions) that shows minors in a sexualized context

Content that depicts child nudity where nudity is defined as

Visible genitalia (even when covered or obscured by transparent clothing)

Visible anus and/or fully nude close-up of buttocks

Uncovered female nipples for children older than toddler-age

No clothes present from neck to knee for children older than toddler-age

Digitally-created depictions of nude minors, unless the image is for health or educational purposes

#### **14. Adult Nudity and Sexual Activity**

#### Policy Rationale

We restrict the display of nudity or sexual activity because some people in our community may be sensitive to this type of content. Additionally, we default to removing sexual imagery to prevent the sharing of non-consensual or underage content. Restrictions on the display of sexual activity also apply to digitally created content unless it is posted for educational, humorous, or satirical purposes.

Our nudity policies have become more nuanced over time. We understand that nudity can be shared for a variety of reasons, including as a form of protest, to raise awareness about a cause, or for educational or medical reasons. Where such intent is clear, we make allowances for the content. For example, while we restrict some images of female breasts that include the nipple, we allow other images, including those depicting acts of protest, women actively engaged in breast-feeding, and photos of post-mastectomy scarring. We also allow photographs of paintings, sculptures, and other art that depicts nude figures.

Do not post:

Images of

- Real nude adults, where nudity is defined as
  - Visible genitalia except in the context of birth giving and after-birth moments or health-related situations (for example, gender confirmation surgery, examination for cancer or disease prevention/assessment)  
Visible anus and/or fully nude close-ups of buttocks unless photoshopped on a public figure  
Uncovered female nipples except in the context of breastfeeding, birth giving and after-birth moments, health-related situations (for example, post-mastectomy, breast cancer awareness or gender confirmation surgery) or an act of protest
- Sexual activity, including
- Sexual intercourse
  - Explicit sexual intercourse, defined as mouth or genitals entering or in contact with another person's genitals or anus, where at least one person's genitals are nude  
Implied sexual intercourse, defined as mouth or genitals entering or in contact with another person's genitals or anus, even when the contact is not directly visible, except in cases of a sexual health context, advertisements, and recognized fictional images or with indicators of fiction  
Implied stimulation of genitalia/anus, defined as stimulating genitalia/anus or inserting objects into genitalia/anus, even when the activity is not directly visible, except in cases of sexual health context, advertisements, and recognized fictional images or with indicators of fiction
  - Other sexual activities including (but not limited to)
  - Erections  
Presence of by-products of sexual activity  
Stimulating genitals or anus, even if above or under clothing  
Use of sex toys, even if above or under clothing  
Stimulation of naked human nipples  
Squeezing female breast except in breastfeeding context  
Fetish content that involves

Acts that are likely to lead to the death of a person or animal

Dismemberment

Cannibalism

Feces, urine, spit, snot, menstruation, or vomit

Digital content that meets our definition of sexual activity unless any of the following conditions exist

Content where the sexual activity (intercourse or other sexual activities) is not directly visible

Content was posted in a satirical or humorous context

Content was posted in an educational or scientific context

Imagery is not sufficiently detailed and only body shapes or contours are visible

## 15. Sexual solicitation

## Policy rationale

As noted in Section 8 of our Community Standards ([Sexual Exploitation of Adults](#)), people use Facebook to discuss and draw attention to sexual violence and exploitation. We recognise the importance of and want to allow for this discussion. We draw the line, however, when content facilitates, encourages or coordinates sexual encounters between adults. We also restrict sexually explicit language that may lead to solicitation because some audiences within our global community may be sensitive to this type of content and it may impede the ability for people to connect with their friends and the broader community.

Do not post:

Content that attempts to coordinate or recruit for adult sexual activities, including but not limited to:

Filmed sexual activities

Pornographic activities, strip club shows, live sex performances, erotic dances

Sexual, erotic or tantric massages

Content that engages in explicit sexual solicitation by, including but not limited to the following, offering or asking for:

Sex or sexual partners

Sex chat or conversations

Nude images

Content that engages in implicit sexual solicitation, which can be identified by offering or asking to engage in a sexual act and/or acts identified by other suggestive elements such as any of the following:

Vague suggestive statements, such as "looking for a good time tonight"

Sexualised slang

Using sexual hints such as mentioning sexual roles, sex positions, fetish scenarios, sexual preference/sexual partner preference, state of arousal, act of sexual intercourse or activity (sexual penetration or self-pleasuring), commonly sexualised areas of the body such as the breasts, groin or buttocks, state of hygiene of genitalia or buttocks

Content (hand-drawn, digital or real-world art) that may depict explicit sexual activity or suggestively posed person(s).

Content that offers or asks for other adult activities such as:

Commercial pornography

Partners who share fetish or sexual interests

Sexually explicit language that adds details and goes beyond mere naming or mentioning of:

A state of sexual arousal (wetness or erection)

An act of sexual intercourse (sexual penetration, self-pleasuring or exercising fetish scenarios)

# Instagram 2019

## Community Guidelines

### The Short

We want Instagram to continue to be an authentic and safe place for inspiration and expression. Help us foster this community. Post only your own photos and videos and always follow the law. Respect everyone on Instagram, don't spam people or post nudity.

### The Long

Instagram is a reflection of our diverse community of cultures, ages, and beliefs. We've spent a lot of time thinking about the different points of view that create a safe and open environment for everyone.

We created the Community Guidelines so you can help us foster and protect this amazing community. By using Instagram, you agree to these guidelines and our Terms of Use. We're committed to these guidelines and we hope you are too. Overstepping these boundaries may result in deleted content, disabled accounts, or other restrictions.

Share only photos and videos that you've taken or have the right to share.

As always, you own the content you post on Instagram. Remember to post authentic content, and don't post anything you've copied or collected from the Internet that you don't have the right to post. Learn more about intellectual property rights.

Post photos and videos that are appropriate for a diverse audience.

We know that there are times when people might want to share nude images that are artistic or creative in nature, but for a variety of reasons, we don't allow nudity on Instagram. This includes photos, videos, and some digitally-created content that show sexual intercourse, genitals, and close-ups of fully-nude buttocks. It also includes some photos of female nipples, but photos of post-mastectomy scarring and women actively breastfeeding are allowed. Nudity in photos of paintings and sculptures is OK, too.

People like to share photos or videos of their children. For safety reasons, there are times when we may remove images that show nude or partially-nude children. Even when this content is shared with good intentions, it could be used by others in unanticipated ways. You can learn more on our Tips for Parents page.

Foster meaningful and genuine interactions.

Help us stay spam-free by not artificially collecting likes, followers, or shares, posting repetitive comments or content, or repeatedly contacting people for commercial purposes without their consent.

Follow the law.

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 illegal or prescription drugs (even if legal in your region) are also not allowed. Instagram also prohibits the sale of live animals between private individuals, though brick-and-mortar stores may offer these sales. No one may coordinate poaching or selling of endangered species or their parts.

Remember to always follow the law when offering to sell or buy other regulated goods. Accounts promoting online gambling, online real money games of skill or online lotteries must get our prior written permission before using any of our products.

We have zero tolerance when it comes to sharing sexual content involving minors or threatening to post intimate images of others.

Respect other members of the Instagram community.

We want to foster a positive, diverse community. We remove content that contains credible threats or hate speech, content that targets private individuals to degrade or shame them, personal information meant to blackmail or harass someone, and repeated unwanted messages. We do generally allow stronger conversation around people who are featured in the news or have a large public audience due to their profession or chosen activities.

It's never OK to encourage violence or attack anyone based on their race, ethnicity, national origin, sex, gender, gender identity, sexual orientation, religious affiliation, disabilities, or diseases. When hate speech is being shared to challenge it or to raise awareness, we may allow it. In those instances, we ask that you express your intent clearly.

Serious threats of harm to public and personal safety aren't allowed. This includes specific threats of physical harm as well as threats of theft, vandalism, and other financial harm. We carefully review reports of threats and consider many things when determining whether a threat is credible.

Maintain our supportive environment by not glorifying self-injury.

The Instagram community cares for each other, and is often a place where people facing difficult issues such as eating disorders, cutting, or other kinds of self-injury come together to create awareness or find support. We try to do our part by providing education in the app and adding information in the Help Center so people can get the help they need.

Encouraging or urging people to embrace self-injury is counter to this environment of support, and we'll remove it or disable accounts if it's reported to us. We may also remove content identifying victims or survivors of self-injury if the content targets them for attack or humor.

Be thoughtful when posting newsworthy events.

We understand that many people use Instagram to share important and newsworthy events. Some of these issues can involve graphic images. Because so many different people and age groups use Instagram, we may remove videos of intense, graphic violence to make sure Instagram stays appropriate for everyone.

We understand that people often share this kind of content to condemn, raise awareness or educate. If you do share content for these reasons, we encourage you to caption your photo with a warning about graphic violence. Sharing graphic images for sadistic pleasure or to glorify violence is never allowed.

Help us keep the community strong:

Each of us is an important part of the Instagram community. If you see something that you think may violate our guidelines, please help us by using our built-in reporting option. We have a global team that reviews these reports and works as quickly as possible to remove content that doesn't meet our guidelines. Even if you or someone you know doesn't have an Instagram account, you can still file a report. When you complete the report, try to provide as much information as possible, such as links, usernames, and descriptions of the content, so we can find and review it quickly. We may remove entire posts if either the imagery or associated captions violate our guidelines.

You may find content you don't like, but doesn't violate the Community Guidelines. If that happens, you can unfollow or block the person who posted it. If there's something you don't like in a comment on one of your posts, you can delete that comment.

Many disputes and misunderstandings can be resolved directly between members of the community. If one of your photos or videos was posted by someone else, you could try commenting on the post and asking the person to take it down. If that doesn't work, you can file a copyright report. If you believe someone is violating your trademark, you can file a trademark report. Don't target the person who posted it by posting screenshots and drawing attention to the situation because that may be classified as harassment.

We may work with law enforcement, including when we believe that there's risk of physical harm or threat to public safety.

For more information, check out our [Help Center](#) and [Terms of Use](#).

Thank you for helping us create one of the best communities in the world,

The Instagram Team

## **Tumblr 2017**

As a global platform for creativity and self-expression, Tumblr is deeply committed to supporting and protecting freedom of speech. At the same time, we draw lines around a few narrowly-defined but deeply important categories of content and behavior that jeopardize our users, threaten our infrastructure, and damage our community.

What Tumblr is for:

Tumblr celebrates creativity. We want you to express yourself freely and use Tumblr to reflect who you are, and what you love, think, and stand for.

What Tumblr is not for:

**Malicious Speech.**

Don't encourage violence or hatred. Don't make violent threats or statements that incite violence, including threatening or promoting terrorism. Especially don't do so on the basis of things like race, ethnic origin, religion, disability, gender, gender identity, age, veteran status, or sexual orientation. If you encounter negative speech that doesn't rise to the level of violence or threats of violence, we encourage you to dismantle negative speech through argument rather than censorship. That said, if you encounter anything especially heinous, tell us about it.

**Harm to Minors.**

Be thoughtful when posting anything involving a minor. Don't post or solicit anything relating to minors that is sexually suggestive or violent. Don't bully minors, even if you are one. Being a teenager is complicated enough without the anxiety, sadness, and isolation caused by bullying.

**Promotion or Glorification of Self-Harm.**

Don't post content that actively promotes or glorifies self-harm. This includes content that urges or encourages others to cut or injure themselves; embrace anorexia, bulimia, or other eating disorders; or commit suicide rather than, e.g., seeking counseling or treatment, or joining together in supportive conversation with those suffering or recovering from depression or other conditions. Dialogue about these behaviors is incredibly important and online communities can be extraordinarily helpful to people struggling with these difficult conditions. We aim for Tumblr to be a place that facilitates awareness, support and recovery, and we will remove only those posts or blogs that cross the line into active promotion or glorification of self-harm.

**Unmarked sensitive content.**

Tumblr is common ground for millions of people from a wide variety of locations, cultures, and backgrounds. Some of them are okay with seeing sensitive content on their dashboard, and some aren't, and that's fine—that's what Safe Mode is for. While we attempt to automatically categorize content that

might not be suitable for all viewers, you can help us respect the choices of people in the community by pre-emptively marking any sensitive material you post. If you're thinking about posting something that contains sensitive content, please flip the "Mark as sensitive" switch before publishing it (you'll see that switch in the post form). Some examples of sensitive content might include nudity in an artistic, educational, or photojournalistic context. If you regularly post sexually explicit material, we ask that you mark your Tumblr as "Explicit" (which you can do on your blog's settings page).

#### Gore, Mutilation, Bestiality, or Necrophilia.

Don't post gore just to be shocking. Don't showcase the mutilation or torture of human beings, animals, or their remains. Dick.

#### Unflagged NSFW Blogs.

Tumblr is home to millions of readers and creators from a variety of locations, cultures, and backgrounds who hold different points of view concerning adult-oriented content. If you regularly post sexual or adult-oriented content, respect the choices of people in our community who would rather not see such content by flagging your blog (which you can do on your blog's Settings page) as Not Suitable for Work ("NSFW"). This action doesn't prevent you or your readers from using any of Tumblr's social features, but rather screens your blog's content from Tumblr users who would prefer not to see NSFW material.

#### Uploading Sexually Explicit Video.

You can embed anything in a Tumblr post as long as it's lawful and it follows our other guidelines, but please don't use Tumblr's Upload Video feature to upload sexually explicit video. We're not in the business of hosting adult-oriented videos (it's fucking expensive).

#### Non-Genuine Social Gesture Schemes.

Don't participate in schemes to drive up non-genuine follows, likes, reblogs, etc. Don't orchestrate or engage in "follow trains," where users are encouraged to follow lists of other users to gain more followers for themselves. Don't make bulk or indiscriminate use of messaging features like Fan Mail or Asks, to, for example, bait reblogs/follows or drive traffic to your blog or website. If you want people to like you, just play it cool and be yourself.

#### Deceptive or Fraudulent Links.

Don't post deceptive or fraudulent links in your posts. This includes giving links misleading descriptions, putting the wrong "source" field in a post, setting misleading click-through links on images, or embedding links to interstitial or pop-up ads.

#### Misattribution or Non-Attribution.

Make sure you always give proper attribution and include full links back to original sources. When you find something awesome on Tumblr, reblog it instead of reposting it. It's less work and more fun, anyway. When reblogging something, DO NOT inject a link back to your blog just to steal attention from the original post.

#### Username / URL Abuse or Squatting.

Tumblr's usernames/URLs are meant for the use and enjoyment of all of our users. Don't squat, hoard, amass, accumulate, accrue, stockpile, rack up, buy, trade, sell, launder, invest in, ingest, get drunk on, cyber with, grope, or jealously guard Tumblr usernames/URLs.

#### Account Dormancy.

Use Tumblr! Use Tumblr all the time! Or at the very least, use Tumblr once a year. If you don't, we may mark your account as dormant. Your content won't go anywhere—it'll be archived exactly as you left it—but your URL(s) will be released for someone else to use.

#### Spam.

Don't spam people. Don't make spammy posts, don't post spammy replies, don't send people spammy messages. Be a regular human. Don't put tags on your posts that will mislead or deceive searchers. For example, don't tag a photo of your cat with "doctor who" unless the name of your cat is actually Doctor Who, and don't overload your posts with #barely #relevant #tags. Of course, meaningful uses of tags are always fine (for example, ironic "punchline" tags that add meaning or context to a post). Don't put dubious code in your posts, like using JavaScript to cause redirects or inject unwanted ads in blogs. Don't use deceptive means to generate revenue or traffic, or create blogs with the primary purpose of affiliate marketing. Spam doesn't belong on Tumblr.

#### Mass Registration or Automation.

Don't register accounts or post content automatically, systematically, or programmatically.

#### Unauthorized Contests, Sweepstakes, or Giveaways.

Please follow our guidelines for contests, sweepstakes, and giveaways.

#### Themes Distributed by Third Parties.

To ensure the safety and stability of our blog network, it's important that themes are distributed through [tumblr.com/themes](http://tumblr.com/themes). It's easy, it's secure, and any interested developer can use it to host and promote their work. Besides the inconvenience involved in using them, third-party theme repositories are often used to inject nefarious code that subjects users to spam, ads, or phishing attacks. Don't use them. Creating blogs or directories that curate themes from [tumblr.com/themes](http://tumblr.com/themes) is a-okay.

#### Copyright or Trademark Infringement.

Respect the copyrights and trademarks of others. If you aren't allowed to use someone else's copyrighted or trademarked work (either by license or by legal exceptions and limitations such as fair use), don't post it. Intellectual property is a tricky issue, so now is as good a time as any to explain some aspects of the process we use for handling copyright and trademark complaints. We respond to notices of alleged copyright infringement as per our Terms of Service and the Digital Millennium Copyright Act; please see our DMCA notification form to file a copyright claim online. Please note that we require a valid DMCA notice before removing content. Parties asserting a trademark infringement claim should identify the allegedly infringing work and the legal basis for their claim, and include the registration and/or application number(s) pertaining to their trademark. Each claim is reviewed by a trained member of our Trust and Safety team.

If we remove material in response to a copyright or trademark claim, the user who posted the allegedly infringing material will be provided with information from the complainant's notice (like identification of the rightsholder and the allegedly infringed work) so they can determine the basis of the claim.

We always want to make sure there is room in any copyright or trademark complaint for both parties to make their case. With regard to copyright claims, the posting user can file a DMCA counter-notification with us, as described in our Terms of Service. Counter-notifications that we determine to be valid will result in restoration of the content at issue following the required waiting period prescribed by the DMCA. With regard to trademark claims, the posting user can send us an appeal explaining their side of the situation, along with any relevant materials we should look at. A successful trademark appeal will also result in restoration of the content at issue.



With regard to repeat copyright infringement, we use a three-strike system to evaluate the standing of a user's account, where, generally, each valid copyright infringement notice constitutes a strike, and three strikes results in the termination of a user's account. When a user submits a valid DMCA counter-notification, we remove the associated strike from their record.

Whew. It's a complex process, but one we're proud of, and that we think strikes all the appropriate balances.

#### Confusion or Impersonation.

Don't do things that would cause confusion between you or your blog and a person or company, like registering a deliberately confusing URL. Don't impersonate anyone. While you're free to ridicule, parody, or marvel at the alien beauty of Benedict Cumberbatch, you can't pretend to actually be Benedict Cumberbatch.

#### Harassment.

Don't engage in targeted abuse or harassment. If anyone is sending you unwanted messages or reblogging your posts in an abusive way, we encourage you to be proactive and block the hell out of them. And if someone blocks you, don't attempt to circumvent the block feature or otherwise try to communicate with them. Just stop.

#### Privacy Violations.

Don't use Tumblr to deceptively obtain personal information. Don't post content that violates anyone's privacy, especially personally identifying or confidential information like credit card numbers, social security numbers, or unlisted contact information. Absolutely do not post non-consensual pornography—that is, private photos or videos taken or posted without the subject's consent.

#### Disruptions, Exploits, or Resource Abuse.

Our servers and the valiant engineers behind them work hard for you. Don't attempt unauthorized use, disruption, or exploitation of Tumblr.com or our other products and services, or otherwise abuse Tumblr's resources.

#### Unlawful Uses or Content.

Don't use Tumblr to conduct illegal behavior, like fraud or phishing. That should be pretty obvious to you, a decent human being.

If we conclude that you are violating these guidelines, you may receive a notice via email. If you don't explain or correct your behavior, we may take action against your account. We do our best to ensure fair outcomes, but in all cases we reserve the right to suspend accounts, or remove content, without notice, for any reason, but particularly to protect our services, infrastructure, users, and community. We reserve the right to enforce, or not enforce, these guidelines in our sole discretion, and these guidelines don't create a duty or contractual obligation for us to act in any particular manner.

You can report violations of these guidelines to us directly.

You can remove your own content whenever you want, of course. If you need help doing that, you can find instructions over in our help docs.

<https://www.tumblr.com/docs/posting#deletpost>

We also reserve the right to amend these guidelines using the procedures set forth in our Terms of Service.

If you have questions or feedback, don't hesitate to let us know.

Thanks for reading all of this, by the way. Welcome to Tumblr.

Link to Prior Versions

You will find prior versions of our Community Guidelines on GitHub, which will allow you to compare historical versions and see which terms have been updated:  
<http://github.com/tumblr/policy/commits/master/community-guidelines.txt>

## Tumblr 2018

Community Guidelines **Last Modified: 2018-12-17**

As a global platform for creativity and self-expression, Tumblr is deeply committed to supporting and protecting freedom of speech. At the same time, we draw lines around a few narrowly defined but deeply important categories of content and behavior that jeopardize our users, threaten our infrastructure, and damage our community.

What Tumblr is for:

Tumblr celebrates creativity. We want you to express yourself freely and use Tumblr to reflect who you are, and what you love, think, and stand for.

What Tumblr is not for:

- **Terrorism.** We don't tolerate content that promotes, encourages, or incites acts of terrorism. That includes content which supports or celebrates terrorist organizations, their leaders, or associated violent activities. **Report terrorism**
- **Hate Speech.** Don't encourage violence or hatred. Don't post content for the purpose of promoting or inciting the hatred of, or dehumanizing, individuals or groups based on race, ethnic or national origin, religion, gender, gender identity, age, veteran status, sexual orientation, disability or disease. If you encounter content that violates our hate speech policies, please report it. **Report hate speech.** Keep in mind that a post might be mean, tasteless, or offensive without necessarily encouraging violence or hatred. In cases like that, you can always block the person who made the post—or, if you're up for it, you can express your concerns to them directly, or use Tumblr to speak up, challenge ideas, raise awareness or generate discussion and debate.
- **Harm to Minors.** Be thoughtful when posting anything involving a minor. Don't post or solicit anything relating to minors that is sexually suggestive or violent. Don't bully minors, even if you are one. Being a teenager is complicated enough without the anxiety, sadness, and isolation caused by bullying. **Report harm to minors**
- **Promotion or Glorification of Self-Harm.** Don't post content that actively promotes or glorifies self-harm. This includes content that urges or encourages others to: cut or injure themselves; embrace anorexia, bulimia, or other eating disorders; or commit suicide rather than, e.g., seeking counseling or treatment, or joining together in supportive conversation with those suffering or recovering from depression or other conditions. Dialogue about these behaviors is incredibly important and online communities can be extraordinarily helpful to people struggling with these difficult conditions. We aim for Tumblr to be a place that facilitates awareness, support and recovery, and we will remove only those posts or blogs that cross the line into active promotion or glorification of self-harm. **Report self-harm**
- **Adult Content.** Don't upload images, videos, or GIFs that show real-life human genitals or female-presenting nipples — this includes content that is so photorealistic that it could be mistaken for featuring real-life humans (nice try, though). Certain types of artistic, educational, newsworthy, or political content featuring nudity are fine. Don't upload any content, including images, videos, GIFs, or illustrations, that

depicts sex acts. For more information about what this guideline prohibits and how to appeal decisions about adult content, check out our help desk.

- **Violent Content and Threats, Gore and Mutilation.** Don't post content which includes violent threats toward individuals or groups - this includes threats of theft, property damage, or financial harm. Don't post violent content or gore just to be shocking. Don't showcase the mutilation or torture of human beings, animals (including bestiality), or their remains. Don't post content that encourages or incites violence, or glorifies acts of violence or the perpetrators. **Report violent threats. Report gore and mutilation.**
- **Non-Genuine Social Gesture Schemes.** Don't participate in schemes to drive up non-genuine Follows, Likes, Reblogs, etc. Don't orchestrate or engage in "follow trains", where users are encouraged to follow lists of other users to gain more followers for themselves. Don't make bulk or indiscriminate use of messaging features, like Fan Mail or Asks, to, for example, bait Reblogs/Follows or drive traffic to your blog or website. If you want people to like you, just play it cool and be yourself.
- **Deceptive or Fraudulent Links.** Don't post deceptive or fraudulent links in your posts. This includes giving links misleading descriptions, putting the wrong "source" field in a post, setting misleading click-through links on images, or embedding links to interstitial or pop-up ads.
- **Misattribution or Non-Attribution.** Make sure you always give proper attribution and include full links back to original sources. When you find something awesome on Tumblr, reblog it instead of reposting it. It's less work and more fun, anyway. When reblogging something, DO NOT inject a link back to your blog just to steal attention from the original post. **Report misattribution or non-attribution**
- **Username/URL Abuse or Squatting.** Tumblr's usernames/URLs are meant for the use and enjoyment of all of our users. Don't squat, hoard, amass, accumulate, accrue, stockpile, rack up, buy, trade, sell, launder, invest in, ingest, get drunk on, cyber with, grope, or jealously guard Tumblr usernames/URLs.
- **Account Dormancy.** Use Tumblr! Use Tumblr all the time! Or at the very least, use Tumblr once a year. If you don't, we may mark your account as dormant. Your content won't go anywhere—it'll be archived exactly as you left it—but your URL(s) will be released for someone else to use.
- **Spam.** Don't spam people. Don't make spammy posts, don't post spammy replies, don't send people spammy messages. Be a regular human. Don't put tags on your posts that will mislead or deceive searchers. For example, don't tag a photo of your cat with "doctor who" unless the name of your cat is actually Doctor Who, and don't overload your posts with #barely #relevant #tags. Of course, meaningful uses of tags are always fine (for example, ironic "punchline" tags that add meaning or context to a post). Don't put dubious code in your posts, like using JavaScript to cause redirects or inject unwanted ads in blogs. Don't use deceptive means to generate revenue or traffic, or create blogs with the primary purpose of affiliate marketing. Spam doesn't belong on Tumblr.
- **Mass Registration or Automation.** Don't register accounts or post content automatically, systematically, or programmatically.
- **Unauthorized Contests, Sweepstakes, or Giveaways.** Please follow our guidelines for contests, sweepstakes, and giveaways.
- **Themes Distributed by Third Parties.** To ensure the safety and stability of our blog network, it's important that themes are distributed through tumblr.com/themes. It's easy, it's secure, and any interested developer can use it to host and promote their work. Besides the inconvenience involved in using them, third-party theme repositories are often used to inject nefarious code that subjects users to spam, ads, or phishing attacks. Don't use them. Creating blogs or directories that curate themes from tumblr.com/themes is a-okay.
- **Copyright or Trademark Infringement.** Respect the copyrights and trademarks of others. If you aren't allowed to use someone else's copyrighted or trademarked work (either by license or by legal exceptions and limitations such as fair use), don't post it.

Intellectual property is a tricky issue, so now is as good a time as any to explain some aspects of the process we use for handling copyright and trademark complaints. We respond to notices of alleged copyright infringement as per our Terms of Service and the Digital Millennium Copyright Act; please see our DMCA notification form to file a copyright claim online. Please note that we require a valid DMCA notice before removing content. Parties asserting a trademark infringement claim should identify the allegedly infringing work and the legal basis for their claim, and include the registration and/or application number(s) pertaining to their trademark. Each claim is reviewed by a trained member of our Trust and Safety team.

If we remove material in response to a copyright or trademark claim, the user who posted the allegedly infringing material will be provided with information from the complainant's notice (like identification of the rightsholder and the allegedly infringed work) so they can determine the basis of the claim.

We always want to make sure there is room in any copyright or trademark complaint for both parties to make their case. With regard to copyright claims, the posting user can file a DMCA counter-notification with us, as described in our Terms of Service. Counter-notifications that we determine to be valid will result in restoration of the content at issue following the required waiting period prescribed by the DMCA. With regard to trademark claims, the posting user can send us an appeal explaining their side of the situation, along with any relevant materials we should look at. A successful trademark appeal will also result in restoration of the content at issue.

With regard to repeat copyright infringement, we use a three-strike system to evaluate the standing of a user's account, where, generally, each valid copyright infringement notice constitutes a strike, and three strikes results in the termination of a user's account. When a user submits a valid DMCA counter-notification, we remove the associated strike from their record.

Whew. It's a complex process, but one we're proud of, and that we think strikes all the appropriate balances. **Report copyright infringement. Report trademark infringement**

- **Confusion or Impersonation.** Don't do things that would cause confusion between you or your blog and a person or company, like registering a deliberately confusing URL. Don't impersonate anyone. While you're free to ridicule, parody, or marvel at the alien beauty of Benedict Cumberbatch, you can't pretend to actually be Benedict Cumberbatch. **Report confusion or impersonation**
- **Harassment.** Don't engage in targeted abuse or harassment. Don't engage in the unwanted sexualization or sexual harassment of others. If anyone is sending you unwanted messages or reblogging your posts in an abusive way, we encourage you to be proactive and block the hell out of them. And if someone blocks you, don't attempt to circumvent the block feature or otherwise try to communicate with them. Just stop. **Report harassment**
- **Privacy Violations.** Don't use Tumblr to deceptively obtain personal information. Don't post content that violates anyone's privacy, especially personally identifying or confidential information like credit card numbers, social security numbers, or unlisted contact information. Absolutely do not post non-consensual pornography—that is, private photos or videos taken or posted without the subject's consent. **Report privacy violations**
- **Disruptions, Exploits, or Resource Abuse.** Our servers, and the valiant engineers behind them, work hard for you. Don't attempt unauthorized use, disruption, or exploitation of Tumblr.com or our other products and services, or otherwise abuse Tumblr's resources.
- **Unlawful Uses or Content.** Don't use Tumblr to conduct illegal behavior, like fraud or phishing. That should be pretty obvious to you, a decent human being.
- **Human Trafficking and Prostitution.** Don't use Tumblr to facilitate sex trafficking, other forms of human trafficking, or illegal prostitution. If you see this activity on Tumblr, report it, and encourage victims to contact law enforcement or the National Human Trafficking Hotline at 1-888-373-7888.

If we conclude that you are violating these guidelines, you may receive a notice via email. If you don't explain or correct your behavior, we may take action against your account. We do our best to ensure fair outcomes, but in all cases we reserve the right to suspend accounts, or remove content, without notice, for any reason, but particularly to protect our services, infrastructure, users, and community. We reserve the right to enforce, or not enforce, these guidelines in our sole discretion, and these guidelines don't create a duty or contractual obligation for us to act in any particular manner.

You can report violations of these guidelines to us directly.

You can remove your own content whenever you want, of course. If you need help doing that, you can find instructions over in our help docs.

We also reserve the right to amend these guidelines using the procedures set forth in our Terms of Service.

If you have questions or feedback, don't hesitate to let us know.

Thanks for reading all of this, by the way. Welcome to Tumblr.

## Twitter 2019

### The Twitter Rules

Twitter's purpose is to serve the public conversation. Violence, harassment and other similar types of behavior discourage people from expressing themselves, and ultimately diminish the value of global public conversation. Our rules are to ensure all people can participate in the public conversation freely and safely.

### Sensitive media policy

#### Overview

#### March 2019

You may not post media that is excessively gory or share violent or adult content within live video or in profile or header images. Media depicting sexual violence and/or assault is also not permitted.

People use Twitter to show what's happening in the world, often sharing images and videos as part of the conversation. Sometimes, this media can depict sensitive topics, including violent and adult content. We recognize that some people may not want to be exposed to sensitive content, which is why we balance allowing people to share this type of media with helping people who want to avoid it to do so.

For this reason, you can't include violent, hateful, or adult content within areas that are highly visible on Twitter, including in live video, profile or header images. If you share this content within Tweets, you need to mark your account as sensitive. Doing so places images and videos behind an interstitial (or warning message), that needs to be acknowledged before your media can be viewed. Using this feature means that people who don't want to see sensitive media can avoid it, or make an informed decision before they choose to view it.

Under this policy, there are also some types of sensitive media content that we don't allow at all, because they have the potential to normalize violence and cause distress to those who view them.

We group sensitive media content into the following categories:

#### Graphic violence

Graphic violence is any media that depicts death, violence, medical procedures, or serious physical injury in graphic detail. Some examples include, but are not limited to, depictions of:

violent crimes or accidents;

physical fights;

physical child abuse;

bodily fluids including blood, feces, semen etc.;

serious physical harm, including visible wounds; and

severely injured or mutilated animals.

Note: exceptions may be made for documentary or educational content.

#### Adult content

Adult content is any consensually produced and distributed media that is pornographic or intended to cause sexual arousal. Some examples include, but are not limited to, depictions of:

full or partial nudity, including close-ups of genitals, buttocks, or breasts (excluding content related to breastfeeding);

simulated sexual acts; and

sexual intercourse or other sexual acts – this also applies to cartoons, hentai, or anime involving humans or depictions of animals with human-like features.

Note: exceptions may be made for artistic, medical, health, or educational content.

For content that was created or distributed without the consent of those featured, please refer to our non-consensual nudity policy.

#### Violent sexual conduct

Violent sexual conduct is any media that depicts violence, whether real or simulated, in association with sexual acts. Some examples include, but are not limited to, depictions of:

rape and other forms of violent sexual assault, or sexual acts that occur without the consent of participants, including a simulated lack of consent; and

sexualized violence – inflicting physical harm on an individual within an intimate setting, where it is not immediately obvious if those involved have consented to take part.

#### Gratuitous gore

Gratuitous gore is any media that depicts excessively graphic or gruesome content related to death, violence or severe physical harm, or violent content that is shared for sadistic purposes. Some examples include, but are not limited to, depictions of:

dismembered or mutilated humans;

charred or burned human remains;

exposed internal organs or bones; and

animal torture or killing. Note: exceptions may be made for religious sacrifice, food preparation or processing, and hunting.

#### Hateful imagery

Hateful imagery is any logo, symbol, or image that has the intention to promote hostility against people on the basis of race, religious affiliation, disability, sexual orientation, gender/gender identity or ethnicity/national origin. Some examples of hateful imagery include, but are not limited to:

symbols historically associated with hate groups, e.g., the Nazi swastika;

images depicting others as less than human, or altered to include hateful symbols, e.g., altering images of individuals to include animalistic features; or

images altered to include hateful symbols or references to a mass murder that targeted a protected category, e.g., manipulating images of individuals to include yellow Star of David badges, in reference to the Holocaust.

What is in violation of this policy?

Our aim is to limit exposure to sensitive images and videos and to prevent the sharing of potentially disturbing types of sensitive media. For this reason, we differentiate our enforcement approach depending on the type of media that has been shared and where it has been shared.

Graphic violence, adult content, and hateful imagery

you can't target people with unsolicited images or videos that contain graphic violence, adult content, or hateful imagery; and

you can't include graphic violence, adult content, or hateful imagery within live video, profile, or header images.

Violent sexual conduct and gratuitous gore

We prohibit violent sexual conduct to prevent the normalization of sexual assault and non-consensual violence associated with sexual acts. We prohibit gratuitous gore content because research has shown that repeated exposure to violent content online may negatively impact an individual's wellbeing. For these reasons, you can't share images or videos that depict violent sexual conduct or gratuitous gore on Twitter. Note: very limited exceptions may be made for gory media associated with newsworthy events.

What is not a violation of this policy?

You can share graphic violence and consensually produced adult content within your Tweets, provided that you mark this media as sensitive. We may also allow limited sharing of hateful imagery, provided that it is not used to promote a terrorist or violent extremist group, that you mark this content as sensitive and don't target it at an individual (via mentioning someone or including an identifiable individual within such images).

To mark your media as sensitive, navigate to your safety settings and select the Mark media you Tweet as containing material that may be sensitive option. If you don't mark your media as sensitive, we will do so manually if your content is reported for review.

Who can report violations of this policy?

Anyone can report potential violations of this policy via our dedicated reporting flows.

How can I report violations of this policy?

In-app

You can report this content for review in-app as follows:

Select Report Tweet from the icon.

Select It displays a sensitive image.

Desktop

You can report this content for review via desktop as follows:

Select Report Tweet from the icon.

Select It displays a sensitive image.

What happens if you violate this policy?

The enforcement action we take depends on the type of media you have shared, and where you have shared it.

Graphic violence, adult content, and hateful imagery

live video and profile images – the first time you violate this policy, we will require you to remove this content. We will also temporarily lock you out of your account before you can Tweet again. If you violate this policy again after your first warning, your account will be permanently suspended;

sending someone unsolicited violent or adult content – if you target someone with sensitive media in a clearly abusive or unsolicited way, we'll require you to remove it under our abusive behavior policy; and

accounts dedicated to posting sensitive media – your account may be permanently suspended if the majority of your activity on Twitter is sharing sensitive media.

If you believe that your account was suspended in error, you can submit an appeal.

Violent sexual conduct and gratuitous gore

if you share media that depicts violent sexual conduct or gratuitous gore anywhere on Twitter, we will require you to remove this content.

if your account is dedicated to posting this type of content, your account will be immediately permanently suspended.

If you believe that your account was suspended in error, you can submit an appeal.

Additional resources

Learn more about our range of enforcement options and our approach to policy development and enforcement.

Adult content that was created or shared without the consent of those depicted is reviewed under our non-consensual nudity policy.

About intimate media on Twitter

**Twitter Rules:** You may not post or share intimate photos or videos of someone that were produced or distributed without their consent. Media depicting sexual violence and/or assault is also not permitted. Note: limited exceptions may apply if there is clear context that the interaction is consensual.

Rationale

Sharing explicit sexual images or videos of someone online without their consent is a violation of their privacy and one of the most serious violations of the Twitter Rules. In addition to posing serious safety and security risks, it may also result in physical, emotional, and financial hardship for the person whose media is shared.

When this applies

We prohibit the posting or sharing of intimate photos or videos that were or appear to have been taken or distributed without the subject's consent. Examples of intimate media that violate this policy include, but may not be limited to:

- hidden camera content involving nudity, partial nudity, and/or sexual acts
- images or videos that appear to have been taken secretly and in a way that allows the user to see the other person's genitals, buttocks, or breasts (content sometimes referred to "creepshots" or "upskirts")
- images or videos that superimpose an individual's face onto another person's nude body;
- images or videos captured in a private setting and not intended for public distribution
- images or videos that are considered and treated as private under applicable laws

Finally, media depicting sexual violence and/or assault (real or simulated) is also not permitted. **Note:** limited exceptions may apply if there is clear context that the interaction is consensual (e.g., consensual sexual violence depicted between actors in television shows or movies).

Do I need to be the target of this content for it to be a violation of the Twitter Rules?



Some forms of adult content and consensual nudity are permitted on Twitter. For this reason, we need to evaluate context to recognize if content has been created and/or shared without the consent of those involved.

We recognize that it is often difficult for those impacted to identify themselves and report this issue. To reduce the burden on those affected, the following types of content **do not require** a first person report;

- creepshots or upskirts - photographs or video taken of people's buttocks, up an individual's skirt/dress or other clothing without consent;
- a bounty or financial reward in exchange for intimate media;
- depictions of sexual violence and/or assault;
- media that is accompanied by;
  - a personal association between the poster and those depicted e.g., "this is my ex"
  - wishing/hoping harm or revenge on those depicted e.g., "I hope that you get what you deserve when people see this"
  - information that could be used to contact those depicted e.g., "you can tell them what you think by calling them on 1234567".

Outside of the above scenarios, we need to hear directly from the individual(s) featured (or their authorized representative) to ensure that we have sufficient context before taking any enforcement action.

#### Consequences

We will immediately and permanently suspend any account that we identify as the original poster of intimate media that was created or distributed without the subject's consent, as well as any account that solely posts this type of content, e.g., accounts dedicated to sharing upskirt images.

In other cases, we may not suspend an account immediately. This is because some people who share this content may do so inadvertently, to express shock, disbelief or to denounce this practice. In these cases, we will require the user to remove the violating content and temporarily lock them out of their account before they can Tweet again. Further violations of this policy will result in permanent account suspension.

Learn more about our range of enforcement options.

If someone believes their account was suspended in error, they can submit an appeal.

#### Child sexual exploitation policy

##### Twitter rules

**Twitter does not tolerate any material that features or promotes child sexual exploitation. This may include media, text, illustrated, or computer generated images.**

##### Rationale

Regardless of the intent, viewing, sharing, or linking to child sexual exploitation media contributes to the re-victimization of the depicted minor and, therefore, is not allowed on Twitter. This also applies to content that may further contribute to victimization of children through promotion or glorification of child sexual exploitation.

##### When this applies

For the purposes of this policy, a minor means any person under the age of 18 years. Examples of content that depicts or promotes child sexual exploitation include, but are not limited to:

- Visual depiction of a minor engaging in sexually explicit or sexually suggestive act.
- Illustrated, computer-generated or other forms of realistic depictions of a human minor in a sexually explicit context, or engaged in a sexually explicit act.
- Links to third-party sites that host child sexual exploitation material.

The following behaviors are also not allowed on Twitter:

- Fantasizing about or promoting engagement in child sexual exploitation.
- Expressing a desire to obtain materials that feature child sexual exploitation.

- Recruiting, advertising or expressing interest in: (1) a commercial sex act involving a child, or (2) in harboring and/or transporting a child for sexual purposes.
- Sending sexually explicit images to a child.
- Engaging a minor in a sexually explicit conversation.
- Trying to obtain sexually explicit images from a minor or trying to recruit a minor for sexual services through incentive or blackmail.
- Identifying alleged victims of childhood sexual exploitation by name or photo.

### Consequences

In the majority of cases, the consequence for violating our child sexual exploitation policy is immediate and permanent suspension. In addition, violators will not be allowed to create new accounts in the future. In a limited number of situations when no malicious intent has been identified, users may instead be asked to remove the violating Tweet and be temporarily locked out of their account before they can Tweet again.

When we are made aware of links depicting or promoting child sexual exploitation they will be removed from the site without further notice and reported to The National Center for Missing & Exploited Children (NCMEC).

### What you can do

If you think you've found a Twitter account distributing or promoting child sexual exploitation, please report it through our designated form. Please provide the username and links to the relevant tweets that lead you to believe the account should be investigated. For help on how to find the direct status links to individual Tweets, see our help page.

If you have found content elsewhere on the internet that is distributing or promoting child sexual exploitation, you should file a report with NCMEC here or with your local law enforcement. If you believe the content is hosted on a website outside of the United States, you can report it on the International Association of Internet Hotlines website.

### Our partners

We partner with multiple organisations around the world whose work is dedicated to fighting child sexual exploitation. You can find a full list of our safety partners here.

## YouTube 2019

### Policies and Safety

When you use YouTube, you join a community of people from all over the world. Every cool, new community feature on YouTube involves a certain level of trust. Millions of users respect that trust and we trust you to be responsible too. Following the guidelines below helps to keep YouTube fun and enjoyable for everyone.

You might not like everything that you see on YouTube. If you think that content is inappropriate, use the flagging feature to submit it for review by our YouTube staff. Our staff carefully reviews flagged content 24 hours a day, 7 days a week to determine whether there's a violation of our Community Guidelines.

### Community Guidelines

### Safety Tools & Resources

### Reporting & Enforcement

Here are some common-sense rules that'll help you steer clear of trouble. Please take these rules seriously and take them to heart. Don't try to look for loopholes or try to lawyer your way around the guidelines—just understand them and try to respect the spirit in which they were created.

### Nudity or sexual content

YouTube is not for pornography or sexually explicit content. If this describes your video, even if it's a video of yourself, don't post it on YouTube. Also, be advised that we work closely with law enforcement and we report child exploitation. **Learn more**

#### Harmful or dangerous content

Don't post videos that encourage others to do things that might cause them to get badly hurt, especially kids. Videos showing such harmful or dangerous acts may be age-restricted or removed depending on their severity. **Learn more**

#### Hateful content

Our products are platforms for free expression. But we don't support content that promotes or condones violence against individuals or groups based on race or ethnic origin, religion, disability, gender, age, nationality, veteran status or sexual orientation/gender identity, or whose primary purpose is inciting hatred on the basis of these core characteristics. This can be a delicate balancing act, but if the primary purpose is to attack a protected group, the content crosses the line. **Learn more**

#### Violent or graphic content

It's not okay to post violent or gory content that's primarily intended to be shocking, sensational or gratuitous. If posting graphic content in a news or documentary context, please be mindful to provide enough information to help people understand what's going on in the video. Don't encourage others to commit specific acts of violence. **Learn more**

#### Harassment and cyberbullying

It's not OK to post abusive videos and comments on YouTube. If harassment crosses the line into a malicious attack it can be reported and may be removed. In other cases, users may be mildly annoying or petty and should be ignored. **Learn more**

#### Spam, misleading metadata and scams

Everyone hates spam. Don't create misleading descriptions, tags, titles or thumbnails in order to increase views. It's not okay to post large amounts of untargeted, unwanted or repetitive content, including comments and private messages. **Learn more**

#### Threats

Things like predatory behaviour, stalking, threats, harassment, intimidation, invading privacy, revealing other people's personal information and inciting others to commit violent acts or to violate the Terms of Use are taken very seriously. Anyone caught doing these things may be permanently banned from YouTube. **Learn more**

#### Copyright

Respect copyright. Only upload videos that you made or that you're authorised to use. This means don't upload videos that you didn't make, or use content in your videos that someone else owns the copyright to, such as music tracks, snippets of copyrighted programs or videos made by other users, without necessary authorisations. Visit our Copyright Centre for more information. **Learn more**

#### Privacy

If someone has posted your personal information or uploaded a video of you without your consent, you can request removal of content based on our Privacy Guidelines. **Learn more**

#### Impersonation

Accounts that are established to impersonate another channel or individual may be removed under our impersonation policy. [Learn more](#)

#### Child safety

Learn about how we protect minors in the YouTube ecosystem. Also, be advised that we work closely with law enforcement and we report child endangerment. [Learn more](#)

If a YouTube creator's on- and/or off-platform behaviour harms our users, community, employees or ecosystem, we may respond based on a number of factors including, but not limited to, the egregiousness of their actions and whether a pattern of harmful behaviour exists. Our response will range from suspending a creator's privileges to account termination.