Moderating online extremism on fringe and mainstream platforms

An analysis of governance by Gab & Twitter

Melissa Blekkenhorst, 4120493

MCMV16048 Master's Thesis NMDC

Utrecht University

Supervisor: Niels Kerssens

Second reader: Mirko Tobias Schäfer

Date: 12-6-2019

Citation: APA Style

Words: 10.028

Abstract

In this research, I have compared the moderation on social media platforms Gab and Twitter. Twitter is a platform that has always promoted free speech but has increasingly strengthened its moderation to prevent harassment and misuse of the platform. Gab was created as an alternative for Twitter and promises free speech and minimal moderation. It is a platform that welcomes users who have been banned from other social media platforms.

An analysis of the guidelines and affordances of Gab and Twitter shows that both of these platforms have had to adjust their moderation strategies to ensure the online safety of the users. Twitter makes an effort to increase the safety in response to user complaints. Gab has had to reinforce its moderation because of complaints from partnering companies that host the site.

After many companies have parted ways with Gab because of controversies and because of the amount of hateful content that the site allows, Gab is looking for ways to host the platform outside of the web infrastructure of the big companies. I argue that the developments of fringe platforms like Gab are a reason for more profound research of the motivations of these platforms and the role that they fulfil within the online public discourse.

Keywords: platform governance, moderation, online extremism, fringe platforms, platform society, online public discourse

List of contents

Introduction

In November 2018, the free speech platform Gab made the decision to ban right-wing politician Patrick Little from its website (Alcorn, 2018). This decision was criticised by users because Gab presents itself as a platform that does not regulate speech and welcomes all opinions. The platform, that was introduced in August 2016 by its founder Andrew Torba, was created as a protest against existing platforms and has the mission to put free speech first (‘Gab’, n.d.). Gab closely resembles the microblogging platform Twitter. Users can create an account, follow other accounts and post short messages on their own timeline. In contrast to Twitter, Gab does not have an extensive list of rules regarding to the content that can be posted:

> We believe that the only valid form of censorship is an individual’s own choice to opt-out. Gab empowers users to filter and remove unwanted followers, words, phrases, and topics they do not want to see in their feeds. However, we do take steps to protect ourselves and our users from illegal activity, spam, and abuse (‘Community Guidelines’, n.d.).

Because of this lack of strict policies Gab attracts many users who have previously been banned from other social media platforms. According to a research by Savvas Zannettou et al. (2018), Gab attracts the interests of users ranging from alt-right supporters and conspiracy theorists to internet trolls. Just like Twitter, the site is mostly used to discuss news, world events and political-related topics (Zannettou et al., 2018). However, the users on Gab share more hateful content than users on Twitter (Zannettou et al., 2018). Because of the amount of hateful content on the platform, Gab has gotten into trouble with multiple organizations who facilitate the use of the site. The app has been deleted from the Google’s Play Store and Apple’s App Store (Price, 2017). Several companies including PayPal ended the collaboration with Gab after the Pittsburgh synagogue shooting in October 2018, because the shooter was active on the platform and had a history of sending antisemitic messages to his followers (Liptak, 2018).

The suspension of Patrick Little’s Gab account indicates that even free speech platforms like Gab have to set boundaries and enforce their own rules in order to stay online. In their book *The Platform Society*, José van Dijck, Thomas Poell and Martijn de Waal (2018) state that each platform has its own way of governing and moderating the content that users post. Most platforms will have a list of rules or guidelines. According to Van Dijck et al. (2018) these guidelines are an important instrument for platforms to govern the relationship between users, partners, clients and other parties. Platform guidelines emerge from a negotiation between platform owners and users, and often change through time (Van Dijck et al., 2018). Together with a platform’s interface, that affords a certain usage, the guidelines shape online discourses. Tarleton Gillespie (2018) states that moderation has become an essential part of governing platforms and argues that platforms should take on the responsibility to act as curators of the online discourse. Both Gillespie and Van Dijck have focused their research on the governance of big platforms like Facebook, Twitter, YouTube, and Tumblr and emphasise that these platforms have an important role in the shaping of online culture. (Gillespie, 2018; Van Dijck et al., 2018).

In my research, I focus on the relationship between the big companies that dominate the platform ecosystem (Van Dijck et al., 2018) and Gab as a fringe platform that is part of a bigger movement that distances itself from this infrastructure. By comparing the governance by Twitter and Gab I will provide insight in how Gab differentiates itself from Twitter. I will focus on how user generated content on Gab and Twitter are moderated through guidelines, technical affordances, and enforcement of the guidelines. Because Gab takes a different approach towards moderating, the site has become a safe space for online extremism. This has not gone unnoticed and has had negative consequences for Gab. However, Gab has managed to keep the website online. My analysis will address the elements of Gab's governance that set it apart from the big tech companies to discuss the role of fringe platforms in the platform ecosystem. I will analyse different aspects of both Twitter and Gab, their sociocultural contexts and the enforcement of the rules by Gab to answer the following research question and sub questions:

How is user generated content on Gab moderated through policy guidelines, technical affordances and enforcement in order to distinguish the platform from Twitter?

1. How have the platform guidelines of Twitter and Gab changed through the years?
2. What is the role of the technical affordances in the moderation of user-generated content on Twitter and Gab?
3. How are the platform guidelines of Gab enforced and how do other parties play a role in the moderation of content on Gab?

1. Theoretical framework: hateful content, platform governance and moderation

Online extremism is not a new phenomenon. In the beginning of the World Wide Web there were already several online communities where extremists would come together and discuss topics of common interest. Mark Dery (1994) wrote a book about 'flame wars', which he describes as verbal fights that would take place on the early internet. When communicating via digital technologies users cannot read each other's body language, so online conversations would often lead to misinterpretations, especially between groups of people who do not agree on something (Dery, 1994). Val Burris, Emery Smith and Ann Strahm (2000) did a research on white supremacist networks on the internet in the 1990's and found several of these hate groups that were active online. Tamara Shepherd, Alison Harvey, Tim Jordan, Sam Srauy and Kate Miltner (2015) notes that online hate is most of the time directed at minorities. Karla Mantilla's (2013) article on 'gendered trolling' is an example of this, as it discusses a form of internet bullying that is specifically aimed at women. Shepherd et al. (2015) states that hate does not only derive from differences between groups, but from power relations between dominant groups and minorities. This notion is important in the context of Gab, since this platform is known for containing expressions of racism, antisemitism, sexism and homophobia (Zannettou et al., 2018). In my analysis, I will discuss the measures that Twitter and Gab have taken to prevent harassment on their platforms.

Danielle Keats Citron (2014) writes about different forms of online harassment in her book *Hate Crimes in Cyberspace*. She makes a distinction between the following forms that are often understood as toxic online behaviours: "threats of violence, privacy invasions, reputation-harming lies, calls for strangers to physically harm victims, and technological attacks" (Citron, 2014, pp. 3). Discriminating speech directed at minorities, or hate speech, is not always treated as content that should be moderated by platforms, because it contains online expressions and not actions (Citron, 2014). Citron (2014, pp. 231) states however that hate speech sets up conditions of violence by framing violence as something that is acceptable when it is directed at certain groups in society. Rodney A. Smolla (1992) makes a distinction between three different ways in which speech can cause harm. First, speech can cause physical harm when it is used to incite others to violence. Second, speech may interfere in social relationships, for example by spreading false information or betraying confidence. And last, speech can cause reactive harm, which includes causing emotional distress and insults to human dignity. These are issues that platforms have to take into account when setting boundaries for types of speech that they permit. My research will show both Twitter and Gab have taken steps to deal with these kinds of harassment.

Tarleton Gillespie (2017) writes about two different kinds of governance in connection to social media: the governance *of* platforms and the governance *by* platforms. When platforms are governed by higher institutions like the national government of a country or for example the European Union, we speak of the governance *of* platforms. The governance *by* platforms is about the ways in which a platform is governed by the owners. In my research, I will mainly focus on this last form of governance. Gillespie (2017) states that platforms impose certain rules upon society by determining what is accepted and what is not. These rules often go further than rules of the law or rules that we deem to be socially acceptable. Online platforms have to take into account that troubling content might scare of advertisers and users. That is why they set up certain guidelines and rules that users have to obey. Gillespie (2017) states that platform owners often have trouble with the enforcement of the rules that they have posed. One of the solutions for this problem is allowing users to help with detecting unacceptable content. With methods like flagging, users can let the platform know when they come across content that is not allowed. But in the end the platform will still decide what is acceptable and what is not. Van Dijck et al. (2018) argues that this is not a democratic process. Users are not in a position to negotiate about platform rules, even though these rules do affect a lot of people. The five biggest platform corporations, Alphabet Inc., Amazon.com Inc., Apple Inc., Facebook Inc. and Microsoft Corporation provide most of the infrastructure that other smaller websites run on (Van Dijck et al., 2018). The rules of these platforms will not only apply to their own websites but are also used as broader guidelines that determine what behaviour is and is not acceptable on the World Wide Web. My analysis will show the impact that this has on the moderation of smaller platforms like Gab.

In his book *Custodians of the Internet* Gillespie (2018) writes about how moderation has become a central part of online platforms. In 1996, the Communications Decency Act made it a criminal act "to display or distribute 'obscene or indecent' material online to anyone under age eighteen" (Gillespie, 2018, pp. 29). Online platforms were granted a 'safe harbor' protection that ensured that they are not responsible for the speech of their users. They are however allowed to intervene and delete content according to their own rules. All platforms are moderated to ensure the online safety of users. However, platforms are not always open and honest about how content is regulated and about how much of the hateful content is actually deleted after being reported. Gillespie (2018) states that examining moderation methods can reveal how a platform works. My research of Gab and Twitter will show that the changes in platform moderation signify a broader change in the way that different platforms enable online discourses.

Gillespie argues that platforms should act as custodians of the internet, which means that they should take the moderation of user-generated content seriously and should provide insight in how this content is moderated: "… not where platforms quietly clean up our mess, but where they take up guardianship of the unresolvable tensions of public discourse, hand back with care the agency for addressing those tensions to users, and responsibly support that process with the necessary tools, data, and insights" (Gillespie, 2018, pp. 216). Furthermore, platforms should collaborate and take on these responsibilities together as an industry-wide commitment (Gillespie, 2018). Gillespie (2018) admits that it might be too much to ask for big platforms to create a consensus on what content should and should not be moderated. I will argue that with the rise of fringe platforms like Gab - that are trying to create an alternative version of the existing platform ecosystem - this issue gets even more complicated.

Adrienne Massanari (2017) argues that online platforms can be places that enable toxic behaviour through lack of supervision and rating systems that can be manipulated by users. Reddit is a platform where users can discuss all kinds of topics by creating message boards called Subreddits. There are only a few rules that Reddit enforces like prohibiting "sharing private information (doxxing), or sexualized images of minors, distributing spam, interfering with the site's regular functioning" (Massanari, 2017, pp. 331). The lack of moderation of content has led to the creation of several Subreddits that include hateful conduct directed at minorities ('reddit', n.d.). Eventually the website decided to ban these Subreddits because of the criticism they received, but an open platform like Reddit still struggles with finding the balance between providing a space for free speech that is also a save space for all users. My research will show that Gab is situated in the similar position of maintaining an open platform while also having to deal with criticisms about hateful content that is posted by its users.

2. The walkthrough method: vision, governance, affordances and enforcement

Because I will analyse multiple elements of Gab and Twitter to explore the governance of these platforms, I have chosen to make use of the walkthrough method as a guideline to my own analysis (Light, Burgess, & Duguay, 2018). The walkthrough method provides a structured way of research. It is grounded in the actor-network theory and examines apps as sociotechnical artefacts (Light et al., 2018). Drawing from the book *The Culture of Connectivity* by Van Dijck (2013), the walkthrough method focuses on the identification of "the technological mechanisms that shape – and are shaped by – the app's cultural, social, political and economic context" (Light et al., 2018, pp. 886). It is a way of researching an app's interface to understand how it guides users towards a certain usage and shapes their experiences. The method helps making explicit the otherwise implicit processes behind an interface. According to Gillespie (2018), platforms tend to be quiet about their moderation and emphasise that they are merely hosting the content with minimal intervention. A walkthrough of the guidelines and affordances of a platform can reveal the underlying ideologies of the producers (Light et al., 2018). The walkthrough method was designed to analyse apps, but the different elements that are studied through the walkthrough method closely resemble the elements that I am analysing on Gab and Twitter, which makes this method also suitable for my research.

The first part of my analysis will be about Gab and Twitter's *vision* and *governance*. *Vision* involves a platform's purpose, target user base and scenarios of use (Light et al., 2018). *Governance* refers to how a platform manages and regulates user activity to fulfil this vision. Examining a platform's vision and governance allows one to understand how platform owners expect users to behave (Light et al., 2018). Gab was created out of dissatisfaction with the strict rules of Twitter and other big social media networks, which means it poses only the strictly necessary rules. In contrast to Gab, Twitter has an extensive list of rules that was lastly updated in October 2018 and explains what sort of content is not acceptable ('The Twitter Rules', n.d.). I will look at the way that Twitter defines and talks about hateful content and then proceed by analysing the way that Gab is differentiating itself from Twitter. I will look at the purpose of Gab as stated on the homepage and in the community guidelines ('Community Guidelines', n.d.). The founder of Gab, Andrew Torba, regularly talks about his opinions on hate speech and the intentions of the platform on his Gab account ('@a', n.d.). I will analyse these posts to gain insight in Gab's vision and governance. For analysing both the guidelines of Twitter and Gab I will use the Wayback Machine. This is a website that is part of the Internet Archive and contains a database of webpages. The Wayback Machine allows access to previous versions of websites, which makes it a useful tool for locating changes on websites (Rogers, 2017). I will use this tool to examine different versions of the guidelines of Twitter and Gab. A small limitation for my research is that the first versions of Twitter's rules are not available. That is why my research on the Twitter guidelines will focus on the rules from 2010 and onwards.

The second part of the research is based on the *technical walkthrough*, which is focussed on the analysis of the technical affordances. Affordances are the possibilities for action. The concept was coined by James J. Gibson (1979) and was later used in the context of technologies by Ian Hutchby (2001). Analysing the technical affordances provides insight in how platforms enable and prevent certain actions. For my research, I will analyse the platform functions on Gab and Twitter that play a role in the moderation. I have located these by accessing both platforms and looking for functions that can help users when they come across content that violates the rules. This includes Gab's report function, mute function and rating system, which I will compare with Twitter's block function, mute function, and the flagging system that has already been researched by Kate Crawford and Tarleton Gillespie (2016). Discussing the platform functions and the actions that they afford, provides insight into the priorities of the platforms in terms of moderation and users' role in the governance of Gab and Twitter. Just like the first part of my research, this will show that Twitter and Gab have a different approach on content moderation. This does however not mean that users will engage with these platforms accordingly. The analysis of user behaviour requires different methods that will not be a part of my research. I will only focus on how the platforms want to be used instead of how they are actually used.

The last part of my research will be about the enforcement of the rules by Gab according to their policies and the ways in which Gab is controlled through the governance of partner companies. This part connects to a part of the walkthrough method called *assessing evidence of unexpected practices*, which is about malpractices that occur on platforms (Light et al., 2018, pp. 895). As Gillespie (2018) states, platforms are not always open and honest about the way they govern. However, platform owners do sometimes justify choices they had to make in relation to malpractices. The official Gab account is occasionally used to respond to criticisms and explain measures that have been taken to improve the platform ('@gab', n.d.). I will analyse this account to look into the controversy surrounding Gab after the Pittsburgh shooting to find out how the company reacted to the criticism it received when it became clear that the shooter was active on the platform (Coaston, 2018). Next to that, I will study a case in which Gab had to ban a user from the platform for breaking the guidelines. The account of Patrick Little, who was posting hateful messages on Gab, was eventually suspended from the platform for encouraging followers to harass private citizens. Gab's statement on this can be found on the official Gab account, so that is what I will use as the data for this last part of my research. It is important to note that this statement is written to justify moderation choices and might therefore not be a good representation of measures that the platform takes to combat malpractices. That is why I will also use secondary sources, namely articles about Gab from various online news websites, to gain more insight in the context of the platform and the way that it is framed by outside sources.

The main objective is to examine Gab as a fringe platform that positions itself as a countermovement against the mainstream platform ecosystem as established by Van Dijck et al. (2018). By analysing the different aspects of Gab's governance in comparison with Twitter, I hope to gain insight in the way that Gab positions itself as a new platform where all users can share their opinions. I will show that Gab uses some of the same tactics of bigger platforms, but also differentiates its platform from others. Which in turn leads to tensions between Gab and partner companies that host the website.

3. Gab’s guidelines & Twitter’s hateful conduct policy

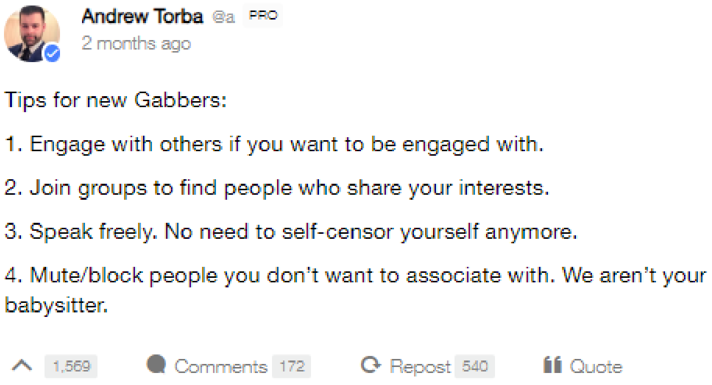
In this first part of the research, I will discuss Gab’s attitude towards moderation and free speech. First, I will discuss Gab’s vision which includes the platform’s purpose, target user base and scenarios of use. Then I will proceed by studying Twitter’s hateful content policy in comparison with Gab’s guidelines to get insight in the platform governance of both platforms.

3.1 The creation of Gab

On the home page of the website, Gab is introduced as: “A social network that champions free speech, individual liberty and the free flow of information online. All are welcome” (‘Gab’, 2016). This implies that the platform wants to attract all sorts of users who value free speech and are looking for an online network were open discussion can take place. Andrew Torba stated in an interview that he felt the need to start his own platform after working in Silicon Valley (Ohlheiser, 2016). According to Torba, most major social networks are operated by progressive leaders. After a Gizmodo article reported that “Facebook workers routinely suppressed news stories of interest to conservative readers from the social network’s influential ‘trending’ news section”, Torba started working on Gab (Nunez, 2016). His new platform was not intended to be a social network for conservatives only, but the site became a safe space for users who are affiliated with the alt-right movement and have previously been suspended from other platforms (Ohlheiser, 2016).

A quote on the frontpage of Gab in 2016 set the tone for the platform: “What is freedom of expression? Without the freedom to offend, it ceases to exist.” This a quote by Salman Rushdie, a British Indian writer who is a vocal advocate for free speech. Rushdie states that free speech is a necessity for a free society. He furthermore argues that speech cannot be free without the risk of offending anyone, because one person’s opinion, might be interpreted as an offensive statement by another (Varshney, 2003). The use of Rushdie’s quote signifies that Gab will not moderate content that might offend other users because this kind of moderation goes in against the principles of free speech.

In December 2018 Torba (‘@a’, n.d.) posted a list of tips for new users. He pinned this list at the top of his timeline, so it is the first post that shows up when one visits his personal account:



Andrew Torba @a PRO
2 months ago

Tips for new Gabbers:

1. Engage with others if you want to be engaged with.

2. Join groups to find people who share your interests.

3. Speak freely. No need to self-censor yourself anymore.

4. Mute/block people you don't want to associate with. We aren't your babysitter.

These tips show that Gab is meant for users who want to engage with other users that share their interests. Users do not have to censor themselves but are given to possibility to mute others if they do not want to interact with them. The last sentence in this post is an indication of what the platform stands for: Gab does not want to interfere in the activities of its users.

Like Gillespie (2018) states, all platforms moderate. Gab is no exception. When signing up to Gab, users have to agree with a few guidelines before they are able to participate. In chapter 3.3, I will take a closer look at these guidelines, but first I will focus on Twitter's policy.

3.2 The Twitter Rules through the years

In contrast to Gab, Twitter has an extensive list of guidelines referring to the moderation of hateful content. At the time of writing, Twitter provides a list of rules and explanations of what type of content is not allowed on the platform. However, these rules were not yet set up when the platform was first introduced in 2006. Twitter did not actively moderate hateful content in the early years:

> Our goal is to provide a service that allows you to discover and receive content from sources that interest you as well as to share your content with others. We respect the ownership of the content that users share and each user is responsible for the content he or she provides. Because of these principles, we do not actively monitor user's content and will not censor user content, except in limited circumstances described below ('The Twitter Rules', n.d.).

This citation is from the Twitter Rules in July 2010 and is the oldest version of the rules that is available at the Wayback Machine. The way that Twitter is framed as a platform in this section is very similar to how Gab is presented now. The platform can be used to share content with others who are interested in the same topics and the role of Twitter itself is minimal. Like on Gab, the owners of Twitter would not interfere, except in 'limited circumstances'. These circumstances include cases of privacy violations, violence, threats, copyright infringement, spam and abuse of the platform or users. In the version from 2010 these rules were not explained in depth. The rules simply told what behaviour is not allowed.

The guidelines on Twitter have seen multiple chances in the past years. In December 2015 Twitter provided the first version of the guidelines that explained the choice for moderating hateful content:

> We believe in freedom of expression and in speaking truth to power, but that means little as an underlying philosophy if voices are silenced because people are afraid to speak up. In order to ensure that people feel safe expressing diverse opinions and beliefs, we do not tolerate behavior that crosses the line into abuse, including behavior that harasses, intimidates, or uses fear to silence another user's voice ('The Twitter Rules', n.d.).

Instead of acting as a neutral platform, Twitter positions itself here as a mediator to ensure that the website is a safe space for everyone to participate. In addition to this explanation the term hateful conduct was introduced:

> Hateful conduct: You may not promote violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, sexual orientation, gender, gender identity, religious affiliation, age, disability, or disease. We also do not allow accounts whose primary purpose is inciting harm towards others on the basis of these categories ('The Twitter Rules', n.d.).

In this section it is made clear that discrimination of minorities is not allowed on Twitter. The platform seems to recognise that hate speech, or hateful conduct, is used as a way to silence people who are part of a minority group. In December 2017 the hateful conduct policy became a separate page of the guidelines with a broader explanation of what the rules entail. In the last version at the time of writing, it becomes even more apparent that there is a shift in the way that Twitter presents itself as a platform. This version includes an extensive description of the company's mission concerning free expression:

> Twitter's mission is to give everyone the power to create and share ideas and information, and to express their opinions and beliefs without barriers. Free expression is a human right – we believe that everyone has a voice, and the right to use it. Our role is to serve the public conversation, which requires representation of a diverse range of perspectives.
>
> We recognise that if people experience abuse on Twitter, it can jeopardize their ability to express themselves. Research has shown that some groups of people are disproportionately targeted with abuse online. This includes; women, people of color, lesbian, gay, bisexual, transgender, queer, intersex, asexual individuals, marginalized and historically underrepresented communities. For those who identity with multiple underrepresented groups, abuse may be more common, more severe in nature and have a higher impact on those targeted.
>
> We are committed to combating abuse motivated by hatred, prejudice or intolerance, particularly abuse that seeks to silence the voices of those who have been historically marginalized ('The Twitter Rules', n.d.).

In this version, Twitter positions itself as a platform whose role it is to serve public conversation. To ensure that everyone can participate in this conversation, the platform has to interfere when users misbehave. Twitter seems to acknowledge that moderation is necessary on a platform with millions of users with diverse opinions.

The changes in guidelines could imply that Twitter has had to react to situations that impacted the platform. Like Van Dijck (2013) states, a site's governance rules are not set in stone, they are a constant target of negotiation. As a platform grows, as new online communities emerge or as controversies arise, platforms have to develop their moderation strategies accordingly (Gillespie, 2018). Several happenings could have played a role in the changing of Twitter's policies, like the Gamergate controversy in 2014 that led to an increase of the online harassment of women who were speaking out against sexism (Massanari, 2014). Controversies like these spark discussions about the role of platforms. Users, media critics and governments are starting to hold big online platforms accountable for the content that gets posted (Gillespie, 2018). Twitter is responding to these criticisms by trying to improve the platform by increasing moderation and providing explanations of the moderation choices that are made. Gillespie's (2018) work suggests that all big platforms have gone through a similar development of increasing moderation and specifying guidelines like the case of Twitter that I have shown in this analysis. In the next section, I will discuss Gab's moderation. Gab takes a different stance towards regulation, but my analysis will also show some similarities between the developments of moderation on Gab in comparison with big platforms.

3.3 The governance by Gab

In contrast to Twitter, Gab does not have specific rules that forbid hateful conduct. The platform started out with a small set of guidelines in 2016 and these guidelines have been altered two times in January 2017 and May 2019. The first version of the guidelines consists of a short text that states that illegal pornography and posting private information of others is strictly forbidden. Furthermore, expressions of violence and terrorism are not allowed:

> We have a zero tolerance policy for violence and terrorism. Users are not allowed to make threats of, or promote, violence of any kind or promote terrorist organizations or agendas. Such users will be instantly removed and the owning account will be dealt with appropriately per the advice of our legal counsel. We may also report the user to local and/or federal law enforcement per the advice of our legal counsel ('Community Guidelines', n.d.).

Gab draws the line at serious threats and promoting violence. This means that as long as users are not explicitly attacking others, they can express any opinion. The first version of the guidelines concludes with the following sentence: "Try to be nice and kind to one another. We're all human". Remarkably, this sentence was removed with the introduction of the second version of the guidelines that applies since 2 January 2017. This newer version contains a more specific description of what kind of content is and is not allowed.

Gab has implemented guidelines that prohibit the sharing of spam and abusing the platform functions. Furthermore, violating copyrights, impersonation and using Gab for illegal transactions has been added to the list of behaviours that are not allowed. Like on Twitter, these guidelines were an addition to the already existing rules. This means that Gab has most likely also run into problems concerning misbehaving users which eventually led them to add new guidelines. In an interview in 2016, Torba was asked about the guidelines and the content that will be removed from the platform (Nash, 2016). He stated: “We expect these guidelines to develop overtime and we will discuss and get feedback on these changes with the community as we scale” (Nash, 2016). This is in line with the development of moderation on platforms that Gillespie (2018) has described. Platforms that have grown bigger have had to make changes to their guidelines because of misbehaviour of their own users. As Gillespie (2018) also states, some platforms learn from the mistakes of other platforms and take over parts of the guidelines to prevent controversies on their own websites. The changes that Gab made in 2017 could therefore also be a result of the platform learning from mistakes that other platforms made in the past.

Where Twitter prohibits all types of harmful speech, Gab draws the line at speech that can lead to physical harm:

> Users are prohibited from calling for the acts of violence against others, promoting or engaging in self-harm, and/or acts of cruelty, threatening language or behaviour that clearly, directly and incontrovertibly infringes on the safety of another user or individual(s). We may also report the user(s) to local and/or federal law enforcement, as per the advice of our legal counsel (‘Community Guidelines’, n.d.).

Even though Gab is known as a free speech platform where users can express all beliefs and opinions, it does warn users that not all behaviour will be accepted. Gab even goes as far as stating that they may report users to law enforcement when the rules are violated. This became clear when Robert Bowers was suspended from Gab after the mass shooting he committed. Gab states that it provided information on Bowers account to the FBI to help the investigation (Coaston, 2018). Even though this incident caused a great controversy for Gab, Torba stated that it would not change anything for the platform and its ideology of providing free speech (Coaston, 2018).

Unlike other platforms, Gab has chosen to use the United States laws as a guideline to their own rules. In May 2019, the platform added a section to the community guidelines that specifies that the rules on Gab that are based on these national laws, also apply to international users (‘Community Guidelines’, n.d.). Gillespie (2016) states that most platforms have set up rules that go beyond national laws. Gab on the other hand, moderates only what is necessary to keep the content on the platform legal. This attitude has caused trouble for the platform in the form of disagreements with collaborating parties. In the last chapter, I will further examine this case.

The current version of the guidelines specifies the way in which Gab enforces the rules that they have set up. Illegal content will be deleted and users can be suspended temporarily or permanently when they are guilty of violating the guidelines. It is not clear whether Gab actively moderates content. The guidelines indicate that a part of the responsibility is shifted to users, because they are asked to tag their Not Safe For Work (NSFW) posts. The default setting hides all content that is tagged as NSFW, but this setting can be switched off to show all content. According to Gillespie (2018) this is a way for platforms to dodge the responsibility of having to make difficult decisions about showing (some types of) nudity or other content that might be seen as offense.

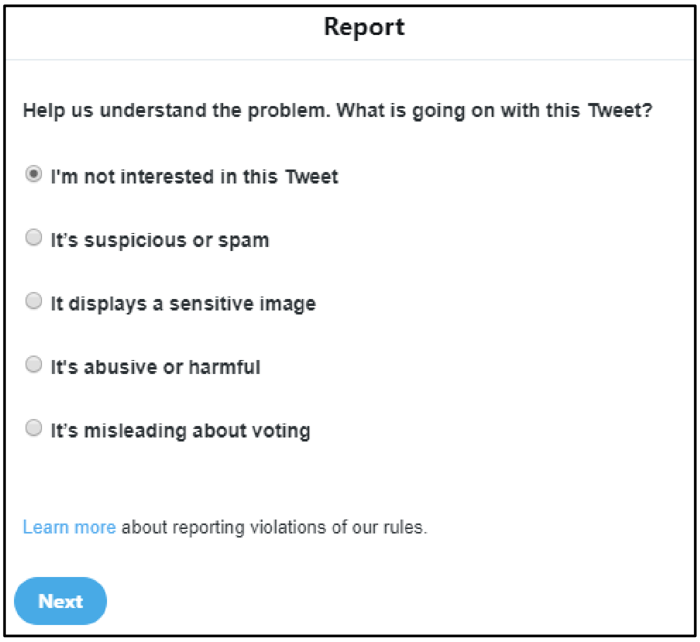
In this chapter, I have shown that there are similarities between the platform guidelines of Gab and Twitter. Even though Gab presents itself as a countermovement against existing platforms and advocates free speech, the guidelines show that moderation and regulation are a necessary part of the platform. However, by asking users to tag their content, the platform avoids having to impose restrictions on NSFW posts. In the next chapter, I will elaborate on other platform functions that allow users to make their own choices about what they want to see in their newsfeed.

4. Twitter & Gab's moderation functions

This chapter is focussed on the platform functions that allow users to secure their own safety on Twitter and Gab. First, I will discuss the flagging, blocking and muting functions on Twitter that have become a part of the platform over the years. Then I analyse the way in which Gab affords users to govern their own accounts. This chapter will show that Twitter has had to make adjustment to the platform in order to ensure the safety of the users. To a lesser extent, the same applies to Gab. Even though Gab has existed for a shorter time span, it has also encountered problems that needed to be solved by making changes in the platform functions.

4.1 Blocking, flagging and muting on Twitter

Just like the guidelines, the affordances of Twitter have changed over time. The addition of extra options for users started in 2008, when Twitter introduced the 'block' function (Twitter Blog, 2008a). According to the platform, this option was requested by users. It allows them to block other users so they cannot be contacted by them anymore. Later that year, Twitter announced additional measures to combat spam accounts (Twitter Blog, 2008b). The company hired people to keep track of accounts that were blocked by many users, and took action to remove them when they turned out to be spam accounts. In 2009, Twitter added the first option that allowed users to report spam (Twitter Blog, 2009). This flagging system has later transformed into an extensive function where not only spam, but also other forms of unaccepted behaviour can be reported. Users can report a tweet via an icon in the corner of every post. The following screen will then appear:



Report

Help us understand the problem. What is going on with this Tweet?

● I'm not interested in this Tweet

○ It's suspicious or spam

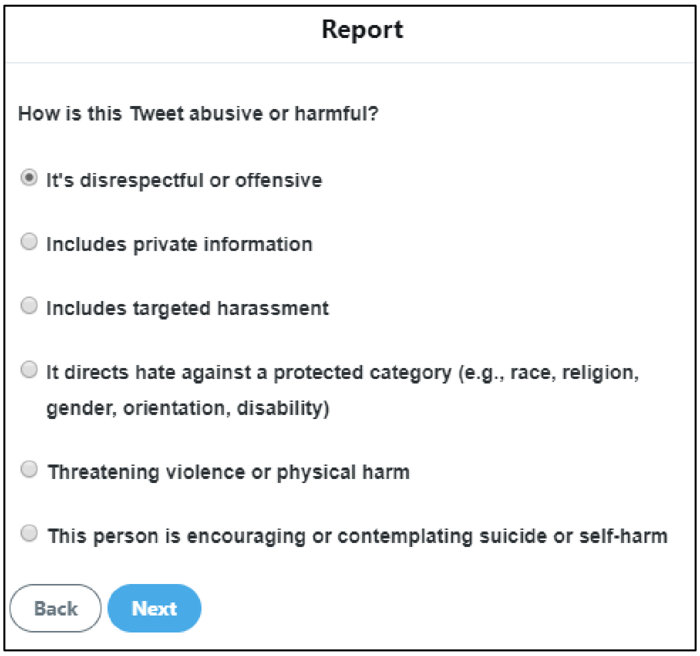
○ It displays a sensitive image

○ It's abusive or harmful

○ It's misleading about voting

Learn more about reporting violations of our rules.

To give Twitter an idea of the kind of content that is being reported, users can choose a category. After submitting, Twitter asks if users want to mute or block the account to ensure that they will not encounter more unacceptable or abusive tweets. In some cases, Twitter will first show a more specific set of options so users can specify the report before it will be submitted. For example, by selecting the option ‘It’s abusive or harmful’ a new screen appears that lets users provide more information about the post:



Report

How is this Tweet abusive or harmful?

● It's disrespectful or offensive

○ Includes private information

○ Includes targeted harassment

○ It directs hate against a protected category (e.g., race, religion, gender, orientation, disability)

○ Threatening violence or physical harm

○ This person is encouraging or contemplating suicide or self-harm

In their article, Crawford and Gillespie (2016) discuss the benefits and implications of flagging systems on social media platforms. They argue that the possibility for reporting offensive content is not only a technical feature, but also “a complex interplay between users and platforms, humans and algorithms, and the social norms and regulatory structures of social media” (Crawford & Gillespie, 2016, pp. 411). In the case of Twitter, the flag option emerged from users’ need to safeguard their own online environment and Twitter’s inability to find and filter out spam accounts. Crawford and Gillespie (2016) state that flagging content is not only a tool that helps the platform locating unallowed behaviour but is also a way to identify users’ opinions about what should and shouldn’t be allowed on social media platforms. By giving users the possibility to specify their reasons for reporting, platforms can learn from their communities and ensure a more user-friendly space (Crawford & Gillespie, 2016). On Twitter there are several options to choose from, but users cannot provide their own explanation for reporting a tweet, which decreases the input of users’ opinions about what should and should not be allowed on the platform. On the other hand, Twitter has made improvements in the communication about flagged posts and accounts. Users now receive updates when Twitter has taken measures against their flagged content, which lets users know that their reports are taken seriously. This change signifies that Twitter is making an effort to improve the transparency of the flagging function to better involve users in the decisions that the platform makes.

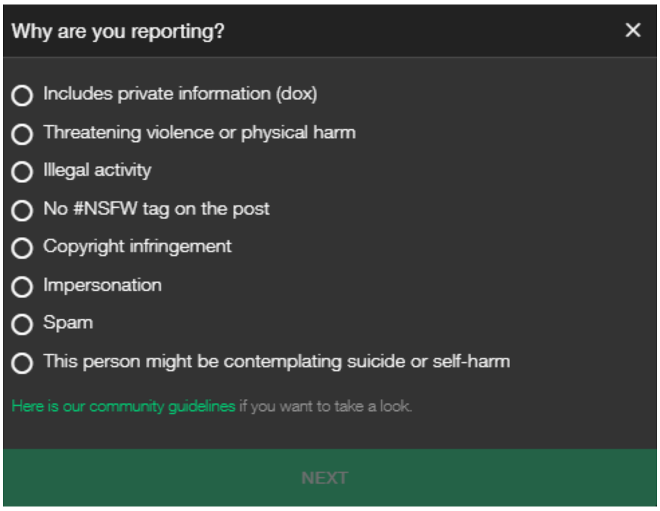
In 2014, the ‘mute’ function was added to Twitter (Twitter Blog, 2014). This option provides a more subtle way of controlling content that users want to interact with. Muting an account hides it from view but unlike the block function the mute is not visible for the account that is muted, which ensures a more anonymous way of controlling content. Next to users, also specific tweets can be muted by entering words, phrases, usernames, emojis and hashtags (Twitter Help Center, n.d.). Tweets that contain the muted content will not be shown on the newsfeed and in notifications.

With the block, report and mute functions that have been developed over time, Twitter now has multiple options for users to ensure their safety on the platform. But despite of the efforts, there is still a demand for improvement. That is why in 2019, Twitter has started proactively reviewing tweets so they do not have to rely on users’ reports (Twitter Blog, 2019). The platform states that a technology is developed that can scan content and notify the moderation team when it encounters questionable content. The development of Twitter functions like these, indicates that the platform has had to adjust the website to the wishes of users. Regardless of whether the modifications of the platform have a positive impact on the user experience, Twitter is taking steps to ensure a safer website for all users by providing more options but also taking on their own responsibility as a platform. This development is not very surprising as Gillespie (2017) states that the governance is an important part of keeping a platform online. In the next section, I will analyse the affordances of Gab that also show a development in platform functions that ensure the online safety of users.

4.2 Up- and downvoting, reporting and muting on Gab

Gab has implemented three different methods that can help users to engage with violating or questionable content. Just like Reddit, Gab has a rating system that allows users to up- and downvote posts (Gab Help, 2018a). Upvoting will increase the visibility of a post and functions as a sign of endorsement of the message. Downvoting will decrease the visibility. The voting system is presumably set up to ensure that the most valuable content, as judged by the users, ends up on top of the page, while unvalued posts are placed lower and form less of a distraction. To make sure that the downvoting option is not misused to silence others, Gab implemented a new system in 2017 that includes taking one point of the users’ total score with every downvote of another’s post (‘@gab’, 2017). The total score is the sum of the number of up- and downvotes a user receives on their posts. A user can only downvote posts with a score of 250 or more, so there is a limited amount of downvotes a user can spend. This limiting of down votes indicates that Gab is, just like Twitter, taking steps to prevent abuse on the website.

The two other options on Gab that allow users to respond to questionable content, are similar to Twitter's functions. The mute option hides content from the muted user from the newsfeed and sends comments from this user to a spam folder (Gab Help, 2018b). Words can also be muted so they will not show up on the newsfeed. Next to mute, Gab also offers a report function. The company does however not provide a lot of information on how this option works. The basic functions are explained in a help section on the site but the report function is not mentioned here (Gab Help, 2018c). The following screen appears when one clicks on the report button that is located in the menu on the corner of every post:



Why are you reporting?

○ Includes private information (dox)
○ Threatening violence or physical harm
○ Illegal activity
○ No #NSFW tag on the post
○ Copyright infringement
○ Impersonation
○ Spam
○ This person might be contemplating suicide or self-harm

Here is our community guidelines if you want to take a look.

NEXT

Similar to Twitter, users on Gab can inform the platform about the kind of content that they wish to report. After making a choice, the website allows users to formulate an explanation in their own words before submitting it. After that, the platform advices to mute or unfollow the user to avoid seeing more content from them. Unlike Twitter, Gab does seem to be interested in the reasons that users have for reporting a post. Another reason for this option, according to Crawford and Gillespie (2016), could be that user complaints can be used to justify governance decisions. When a social media platform decides to delete posts or suspend an account, it can refer to the criticism that the content received to explain their decision to others. In the next chapter I will explore the way that Gab talks about their decisions for suspending users, to find out if these decisions are based upon users' criticisms or motivated by something else.

Even though Gab provides a few options for guaranteeing users' safety, the platform only actively encourages the use of the mute function. The report function is present, but Gab does not provide an explanation of how it works. Crawford and Gillespie (2016) state that many platforms do not inform users on what happens with the reports. This is also the case on Gab. The platform asks users to take responsibly for their own content and only interferes in extreme cases. In the next chapter, I will discuss these cases to get insight in the enforcement of the rules on Gab.

5. Enforcement by & of Gab

In the last two chapters, I have shown that Gab has in some ways utilised the same moderation strategies as the similar platform Twitter. In this chapter, I will elaborate on the ways in which Gab differs from big platforms in terms of moderation and the consequences that this has for Gab. I discuss the enforcement of the guidelines on Gab by analysing a case. Next to that, I will examine the consequences of lack of platform moderation. Partnering companies are not always satisfied with the moderation of Gab, the platform has been banned by multiple partner websites. At the end of this chapter, I discuss a few measures that Gab has taken to avoid being controlled by other authorities.

5.1 Enforcement by Gab

In November 2018, Gab made the decision to suspend Patrick Little's account. In a statement, Gab makes clear that Little's account overstepped the boundaries of the platform. Gab draws a line between hate speech and targeted harassment. Even though Little's account contained language that could be categorised a hate speech, that is not what caused the suspension according to a statement on the official Gab account: "Gab does not moderate speech which is merely offensive or outrageous, including hate speech, as the U.S. Supreme Court has ruled unanimously that 'hate speech' is protected by the First Amendment" ('@gab', 2018). The reason for the suspension is his involvement in targeted harassment:

> In this instance, Gab determined that Gabs posted to the @Patrick_Little account crossed the line by encouraging its followers to harass private citizens who were not the subject of any public controversy. This attempted harassment included but was not limited to offering to distribute the "dox," or personal contact information, of the targets of the harassment. Per a recent appeal of a pre-trial motion to dismiss in Gersh v. Anglin, "exploiting the prejudices widely held among [a publisher's] readers to specifically target" private citizens is not protected by the First Amendment and violates our Community Guidelines.

The highest rated comments on the statement suggest that users were not happy with Gab's decision. Users criticise the platform for not providing any evidence and limiting the free speech ideals that were promised. Ariadna Matamoros-Fernández (2017) argues that arbitrary enforcement of rules creates a platform that reproduces inequality. Policies should be clear and the enforcement or rules should be consistent for all users (Matamoros-Fernández, 2017). Gab makes clear distinctions in its communication about policies and provides elaborate information on what the rules are based on. But judging from this specific case and the users' reaction, Gab does not present any evidence for the suspension of Little's account.

After his suspension from Gab, Little uploaded a YouTube video where he reacts to the way that Gab has treated him. At the time of writing, this video is not available anymore, but an article on Mic.com has covered the events up until Little's suspension from Gab and his reaction afterwards (Alcorn, 2018). In his YouTube video, Little denied doxing anyone on Gab and stated that he was not provided any information about the posts that led to his removal from the site (Alcorn, 2018). However, the article mentions that Little has faced criticism from Gab before because of his antisemitic sentiments, indicating that his removal from the platform did not come unexpected (Alcorn, 2018).

In August 2018, Gab has come under scrutiny from hosting provider Microsoft because of two of Little's Gab posts calling for the 'complete eradication of all Jews' (Gault, 2018). Microsoft demanded for these posts to be taken down as they were a violation of the Microsoft Azure Acceptable Use Policy. Otherwise, Gab would face suspension from Microsoft's hosting services, which means that the platform would be taken offline until it finds a new hosting provider (Gault, 2018). Eventually, Gab complied and made sure that Little's posts would be removed, acknowledging that one of the posts was also a violation of Gab's own guidelines ('Gab user's anti-Semitic posts removed', 2018).

This incident raises questions about the enforcement of the guidelines by Gab. The platform only took action after a warning from Microsoft. Besides, Torba seemed reluctant in doing so, stating: "We are actively looking into other hosting providers and our long-term goal is building our own infrastructure" ('Gab user's anti-Semitic posts removed', 2018). Gab has put out a statement about the final decision to remove Little's account, but it remains unclear what exact happenings led to this.

On a platform where the owners and many users are in favour of free speech, it is more likely that posts like Little's will go unreported, because it is not viewed as troubling content. This means that hateful content will go unmoderated unless other companies like Microsoft interfere. Gillespie (2018) describes a similar situation that occurred on Reddit. On this platform, users can create their own Subreddit, a message board that can be used to discuss all kinds of topics. The Subreddits are moderated by the creators. Just like Gab, Reddit advices to only follow Subreddits of your interests so you will not come into contact with content you do not want to see. But when Subreddits with hateful and illegal content were created and went unmoderated by their creators, the platform intervened and started banning these Subreddits, posed stricter rules and pressed users to moderate more (Gillespie, 2018). Whether or not Gab will ever intervene in this manner is questionable. Just like the Subreddits that were unmoderated because the creators did not view the content as troubling, Gab is likely a safe space for extremist views in a similar way because the creators and other users are not bothered by this kind of content.

5.2 Enforcement of Gab

Several data analyses on Gab show that the platform contains large amounts of hateful content and political extremism (Lima et al., 2018; Mathew, Dutt, Goyal, & Mukherjee, 2018; Zannettou et al., 2018; Zhou, Dredze, Broniatowski, & Adler, 2018). Even though this type of content is accepted on Gab, it has caused trouble for the collaboration with other companies. In the previous section, I have already discussed how Gab had to remove posts because they were in violation of Microsoft's policy. This was however not the first time that Gab has gotten into trouble.

In 2017, both Google and Apple banned the Gab app from their app stores for containing hateful speech that is in violation with their guidelines, and in the case of Apple also for allowing pornography. (Lee, 2017; Nash, 2017). Domain registrar AsiaRegistry ordered Gab to take down a post that mocked Heather Heyer, a woman who was killed during the Unite the Right rally in Charlottesville (Robertson, 2017). In 2018, the Pittsburgh synagogue shooting caused a controversy for Gab because the alleged shooter, Robert Bowers, had been active on the platform (Ohlheiser & Shapira, 2018). Multiple companies broke ties with Gab after this incident, like payment system PayPal, domain registrar GoDaddy and online publishing platform Medium (Ohlheiser & Shapira, 2018). Gab announced on Twitter that hosting provider Joyent had also decided to suspend the platform, which resulted in Gab being offline for a week ('@getongab', 2018).

These examples show that Gab has faced a lot of resistance. It also indicates that small platforms like Gab are very dependent on other companies. Like Van Dijck et al. (2018) argues, a few big corporations have formed an online ecosystem upon which other platforms and apps can be built. The case of Gab shows that the governance of these big companies sets boundaries for the smaller platforms. In the next section, I will discuss the ways in which Gab and other fringe platforms are trying to dodge these restrictions.

5.3 The future of the fringe internet

After Gab was offline for a while in the aftermath of the Pittsburgh shooting, the platform was brought back by domain registrar Epik (Monster, 2018). Epik's CEO Rob Monster states that it was not an easy decision to accept this domain registration after the controversy around Gab. But he welcomed Gab nonetheless because he supports free speech and believes that de-platforming is a form of digital censorship (Monster, 2018). Finding other companies with the same convictions about free speech, is one of the strategies that helped Gab stay online after facing controversy.

According to an article on Slate in 2017, alt-right supporters have become a lot less welcome on the web since the Charlottesville rally (Glaser, 2017). This has resulted in a wish to build an alternative version of the web were the alt-right movement is not censored. At the time of the Charlottesville rally, Gab had already made large steps to becoming an alternative for Twitter with more than 240.000 users and a one million dollar budget raised through crowdfunding (Glaser, 2017). Around the time of the Pittsburgh shooting, Gab had allegedly between 465.000 and 800.000 users (Coaston, 2018). After many companies parted ways with Gab, the platform had to keep finding new ways to stay online. In doing so, Gab and Androw Torba have become frontrunners of what is often referred to as the 'alt-tech', a movement towards creating a web outside of the main infrastructure which is now mostly dominated by Silicon Valley companies (Glaser, 2017).

After being banned from PayPal, Gab started exploring the use of cryptocurrency. Torba states in an interview that he sees Bitcoin as a cryptocurrency that offers censorship resistant payment processing, referring to Bitcoin as free speech money (McCormack, 2019). However, the company Coinbase, which acted as Gab's cryptocurrency payment processor, cut ties with Gab over concerns that the platform encourages hate speech (Owen, 2019). After also being banned from another payment processor, Square's Cash App, Gab started using the open-source server BTCPay that allows them to become their own payment processor (McCormack, 2019). Gab relies financially on GabPro members who pay a monthly or yearly fee in exchange for extra functions like creating groups for users with similar interests and bookmarking posts in order to save and categorise them. The open-source payment system has helped to keep the platform online and according to a Gab post from May 2018, this use of open-source products will be expanded in the future to ensure the existence of Gab as an independent free speech platform ('@gab', 2018).

In the last few years, several other platforms have arisen to function as alternatives for mainstream platforms. A few examples are: WrongThink (alt-Facebook), PewTube (alt-YouTube), Voat (alt-Reddit), Infogalactic (alt-Wikipedia), Hatreon (alt-Patreon) and GoyFundMe (alt-Kickstarter) (Roose, 2017). These platforms struggle with the same problems that Gab is facing. Not only staying out of legal trouble, but also becoming profitable is challenging (Roose, 2017). Gab has managed to raise money through crowdfunding and subscriptions from GabPro members, but some of the other right-extremist platforms have already permanently been taken of the web. However, the rise of these fringe platforms does indicate a resistance against the mainstream online media and the boundaries that they form for individuals with extremist views.

Even though big platforms are putting more and more effort in banning extremism from their websites, they still face a lot of criticism for not doing enough (Koebler, Mead, & Drummond, 2019). An article on Vice states that Twitter is now researching the effects of white supremacism and is questioning whether or not it is the right approach to de-platform individuals with extremist ideas: "Vijaya Gadde, Twitter's head of trust and safety, legal and public policy, said Twitter believes 'counter-speech and conversation are a force for good, and they can act as a basis for de-radicalization, and we've seen that happen on other platforms, anecdotally'" (Koebler et al., 2019). This quote suggests that keeping extremists on the platform will help them de-radicalise. Twitter has considered banning all extremist content, which will create a safer online community, but is a complicated task that takes much effort to do right. Besides, there is the question of where users with extremist convictions will go after they have been banned (Koebler et al., 2019). In the expanding of alternative platforms like Gab, where these people are welcomed, lies the risk of the radicalisation of communities who are already skewed towards extremist ideologies (‘The New Radicalization of the Internet’, 2018). Both Gab and Twitter struggle with the responsibility of hosting a platform that promotes free speech, while also having to pose restrictions to prevent extremism from getting out of hand. It is a complicated situation with conflicting interests for both platforms, that do not only want to stick to their own ideologies and intentions for the platform, but also have to find a way to keep users, advertisers, governments and partnering companies satisfied.

Conclusion

In this research, I have compared the content moderation on Twitter with the moderation of fringe platform Gab. Gab serves as an alternative for Twitter and has attracted the attention of users who have previously been banned from other mainstream platforms for breaking their rules. To analyse the moderation of both platforms, I have focussed on three aspects of platform governance: guidelines, affordances and enforcement. Making use of the walkthrough method (Light et al., 2018), I have analysed the moderation strategies of both platforms in connection to the underlying ideologies that they convey.

Even though Gab was introduced as a platform that allows free speech, there are a few basic restrictions that are posed to avoid legal trouble. Twitter started out in a similar way but had to adjust their guidelines to prevent harassment and misuse of the platform, which resulted in an extensive list of rules. Gab has adjusted its rules twice. The guidelines now express more specific restrictions but compared to Twitter this platform does not have a strict moderation strategy. The main difference between the guidelines of Gab and Twitter is their policy on hate speech. Gab allows hate speech, because this kind of speech is protected by the First Amendment. Twitter forbids hate speech because it is harmful for individuals who are part of a minority group and face discrimination. Gab recommends its users to use the platform only to share their interests with others and to mute other users and content that they do not want to interact with, thereby avoiding harassment.

On both Twitter and Gab there are block and report functions for the protection of users' safety. The development in Twitter's functions shows that the platform had to be improved to make it a safer space. Gab also shows signs of the platform having to respond to misuse, when the rating system was used to silence others' opinions. In comparison, Twitter puts a lot more effort in ensuring a safe online environment for its users. Gab provides a few options but only actively encourages the use of the mute function. The results do however indicate that Gab, both in terms of guidelines and affordances, show some similarities to Twitter. Gab, just like Twitter, has had to adjust its moderation to keep the platform operational. But there are some parts of the platform Gab refused to change. The last part of my research shows that this has led to trouble for Gab.

The lack of strict moderation and the allowance of hate speech were a reason for many companies to part ways with Gab. The case of Gab shows that smaller platforms heavily rely on the online infrastructure that has been created by the bigger companies that Van Dijck et al. (2018) writes about. These bigger companies are in a position to define what is and what is not acceptable in online public discourse, and their governance effects and sets up boundaries for Gab and other fringe platforms. Gab, however, is not discouraged by this. Where Gab has distanced itself from mainstream platforms in terms of ideology, it is now also working to set itself apart technologically by using open-source products that allows the platform to be independent from the bigger platform ecosystem.

On the one hand, this development signifies a change that could have a positive effect on the freedom of the internet and could help smaller platforms get independence. On the other hand, the big platforms that are now providing protection against extremism within the online public discourse could lose their position of power over the public discourse. At the end of his book, Gillespie (2018) argues that platforms, as shapers of public discourse, should take on the responsibility of questioning what is acceptable in online communication and what is not, and should, helped by input from users, act as custodians of the internet. A question that arises, is whether the big platforms will stay in that position, when fringe platforms are started to create their own spaces on the internet.

At this moment of time, it does not look like the fringe platforms will be creating their own alternative version of the web any time soon. And if they do, it remains to be seen whether users are interested enough to switch from mainstream platforms to the alternatives. However, as these fringe platforms grow bigger and more independent, it becomes more relevant to look into the motivations of these platforms and the role that they fulfil within the online public discourse. Especially in the current political landscape, which is becoming more and more polarised, it is important to gain a better understanding of platforms like Gab, that are politically and ideologically motivated to take a stand against bigger platforms that are dominating the internet right now.

Discussion

To analyse the moderation of Twitter and Gab, I have had to narrow down the elements that I could analyse on both sites. I have chosen to focus only on the guidelines and affordances that play a role in the moderation. These elements do possibly not give a representative overview of the entire governance of and by the platforms. The walkthrough method provides instructions on how to gather data from an app, which I have partly made use of. However, I have also left some parts of this method out to narrow down my own analysis. I have not focused on the registration mechanisms of Twitter and Gab and neither on how both platforms are seeking to keep users engaged.

Besides, I have mostly focused on the way that the platforms present themselves. I have deliberately left out the users' perspective, because this was outside of the scope of my research. Additional research about the users' opinions of both sites could give more insight in the role that the platforms serve as spaces that enable online public discourse. Gab has a different user base than Twitter and this likely effects the way that the platforms are governed. This became clear when looking at the reasons for moderation both platforms, which are more internal for Twitter, where users have complained about their safety, as opposed to the external factors like other companies' complaints that played a role in Gab's moderation. A research of the users and their opinions about the platforms can provide more information on the governance and moderation from a user's perspective.

Finally, when I claim that Gab does not moderate unless it has to, I cannot say this with certainty because I only analysed the case of Patrick Little. Based on what happened with Little's Gab account, it seems like Gab was forced to remove him even though the platform and its users might not agree with this decision. In additional research, multiple cases of people who have been banned from Gab could be the object of research to find out more about the underlying reasons of the removals.

Bibliography

Alcorn, C. (2018). Gab just booted white supremacist Patrick Little for anti-Semitic threats. Retrieved 13 March 2019, from https://mic.com/articles/192595/gab-just-booted-white-supremacist-patrick-little-for-anti-semitic-threats

Burris, V., Smith, E., & Strahm, A. (2000). White Supremacist Networks on the Internet. *Sociological Focus*, *33*(2), 215–235. https://doi.org/10.1080/00380237.2000.10571166

Citron, D. K. (2014). *Hate Crimes in Cyberspace*. Cambridge, Massachusetts & London: Harvard University Press.

Coaston, J. (2018, October 29). Gab, the social media platform favored by the alleged Pittsburgh shooter, explained. Retrieved 2 February 2019, from https://www.vox.com/policy-and-politics/2018/10/29/18033006/gab-social-media-anti-semitism-neo-nazis-twitter-facebook

Crawford, K., & Gillespie, T. (2016). What is a flag for? Social media reporting tools and the vocabulary of complaint. *New Media & Society*, *18*(3), 410–428. https://doi.org/10.1177/1461444814543163

Dery, M. (1994). *Flame Wars: The Discourse of Cyberculture*. Durham & London: Duke University Press.

Gab user's anti-Semitic posts removed. (2018). Retrieved 31 January 2019, from BBC News website: https://www.bbc.com/news/technology-45141871

Gault, M. (2018). Microsoft Demands That Gab Delete Post Calling for 'Eradication of All Jews'— Motherboard. Retrieved 31 January 2019, from https://motherboard.vice.com/en_us/article/j5naby/microsoft-demands-that-gab-delete-post-calling-for-eradication-of-all-jews

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. New York: Psychology Press. https://doi.org/10.4324/9781315740218

Gillespie, T. (2017). Governance of and by platforms. *Sage Handbook of Social Media*. London: Sage, 254-278.

Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven & London: Yale University Press.

Glaser, A. (2017, August 30). Nazis and White Supremacists Are No Longer Welcome on the Internet. So They're Building Their Own. Retrieved 30 May 2019, from Slate Magazine website: https://slate.com/technology/2017/08/the-alt-right-wants-to-build-its-own-internet.html

Hutchby, I. (2001). Technologies, Texts and Affordances. *Sociology*, *35*(2), 441–456. https://doi.org/10.1177/S0038038501000219

Koebler, J., Mead, D., & Drummond, K. (2019, May 29). Twitter Has Started Researching Whether White Supremacists Belong on Twitter. Retrieved 31 May 2019, from Vice website: https://www.vice.com/en_us/article/ywy5nx/twitter-researching-white-supremacism-nationalism-ban-deplatform

Lee, T. B. (2017, August 18). Google explains why it banned the app for Gab, a right-wing Twitter rival. Retrieved 29 May 2019, from Ars Technica website: https://arstechnica.com/tech-policy/2017/08/gab-the-right-wing-twitter-rival-just-got-its-app-banned-by-google/

Light, B., Burgess, J., & Duguay, S. (2018). The walkthrough method: An approach to the study of apps. *New Media & Society*, *20*(3), 881–900. https://doi.org/10.1177/1461444816675438

Lima, L., Reis, J. C. S., Melo, P., Murai, F., Araujo, L., Vikatos, P., & Benevenuto, F. (2018). Inside the Right-Leaning Echo Chambers: Characterizing Gab, an Unmoderated Social System. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 515–522. https://doi.org/10.1109/ASONAM.2018.8508809

Liptak, A. (2018, October 27). Paypal bans Gab following Pittsburgh shooting. Retrieved 12 February 2019, from https://www.theverge.com/2018/10/27/18032930/paypal-banned-gab-following-pittsburgh-shooting

Mantilla, K. (2013). Gendertrolling: Misogyny Adapts to New Media. *Feminist Studies*, *39*(2), 563–570.

Massanari, A. (2017). #Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. *New Media & Society*, *19*(3), 329–346. https://doi.org/10.1177/1461444815608807

Matamoros-Fernández, A. (2017). Platformed racism: The mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube. *Information, Communication & Society*, *20*(6), 930–946. https://doi.org/10.1080/1369118X.2017.1293130

Mathew, B., Dutt, R., Goyal, P., & Mukherjee, A. (2018). Spread of hate speech in online social media. *ArXiv:1812.01693 [Cs]*. Retrieved from http://arxiv.org/abs/1812.01693

McCormack, P. (2019, February 10). Gab's Andrew Torba on Why Bitcoin Is Free Speech Money. Retrieved 30 May 2019, from Hacker Noon website: https://hackernoon.com/gabs-andrew-torba-on-why-bitcoin-is-free-speech-money-dcbe15be5e43

Monster, R. (2018). Why Epik welcomed Gab.com. Retrieved 30 May 2019, from Epik Blog website: https://epik.com/blog/why-epik-welcomed-gab-com.html

Nash, C. (2016, August 23). Meet the CEO of Gab, The Free Speech Alternative to Twitter. Retrieved 16 May 2019, from Breitbart website: https://www.breitbart.com/tech/2016/08/23/meet-the-ceo-of-gab-the-free-speech-alternative-to-twitter/

Nash, C. (2017, January 23). Apple App Store Rejects Free-Speech Twitter Alternative 'Gab' AGAIN, Blames User Content. Retrieved 29 May 2019, from Breitbart website:

https://www.breitbart.com/tech/2017/01/22/apple-app-store-rejects-free-speech-twitter-alternative-gab/

Nunez, M. (2016). Former Facebook Workers: We Routinely Suppressed Conservative News. Retrieved 7 April 2019, from https://gizmodo.com/former-facebook-workers-we-routinely-suppressed-conser-1775461006

Ohlheiser, A. (2016). Banned from Twitter? This site promises you can say whatever you want. Retrieved 7 April 2019, from https://www.washingtonpost.com/news/the-intersect/wp/2016/11/29/banned-from-twitter-this-site-promises-you-can-say-whatever-you-want/

Ohlheiser, A., & Shapira, I. (2018). Gab, the white supremacist sanctuary linked to the Pittsburgh suspect, goes offline (for now). Retrieved 15 May 2019, from Washington Post website: https://www.washingtonpost.com/technology/2018/10/28/how-gab-became-white-supremacist-sanctuary-before-it-was-linked-pittsburgh-suspect/

Owen, T. (2019, January 23). Gab is back in business after finding a payments processor willing to work with the alt-right. Retrieved 30 May 2019, from Vice News website: https://news.vice.com/en_us/article/eve43n/gab-is-back-in-business-after-finding-a-payments-processor-willing-to-work-with-the-alt-right

Price, R. (2017, August 18). Google’s app store has banned Gab—A social network popular with the far-right—For ‘hate speech’. Retrieved 1 April 2019, from http://uk.businessinsider.com/google-app-store-gab-ban-hate-speech-2017-8

Robertson, A. (2017, September 6). The far-right’s favorite social network is facing its own censorship controversy. Retrieved 29 May 2019, from The Verge website: https://www.theverge.com/2017/9/6/16259150/gab-ai-registrar-andrew-anglin-daily-stormer-crackdown

Rogers, R. (2017). Doing Web history with the Internet Archive: Screencast documentaries. *Internet Histories*, *1*(1–2), 160–172. https://doi.org/10.1080/24701475.2017.1307542

Roose, K. (2017). The Alt-Right Created a Parallel Internet. It’s an Unholy Mess. *The New York Times*. Retrieved from https://www.nytimes.com/2017/12/11/technology/alt-right-internet.html

Shepherd, T., Harvey, A., Jordan, T., Srauy, S., & Miltner, K. (2015). Histories of Hating. *Social Media + Society*, *1*(2). https://doi.org/10.1177/2056305115603997

Smolla, R. A. (1992). *Free speech in an open society*. New York: Knopf.

The New Radicalization of the Internet. (2018). *The New York Times*. Retrieved from https://www.nytimes.com/2018/11/24/opinion/sunday/facebook-twitter-terrorism-extremism.html

Van Dijck, J. (2013). *The Culture of Connectivity: A Critical History of Social Media*. New York: OUP USA.

Van Dijck, J., Poell, T., & De Waal, M. (2018). *The Platform Society: Public Values in a Connective World*. New York: Oxford University Press.

Zannettou, S., Bradlyn, B., De Cristofaro, E., Kwak, H., Sirivianos, M., Stringhini, G., & Blackburn, J. (2018). What is Gab? A Bastion of Free Speech or an Alt-Right Echo Chamber? *Companion of the The Web Conference 2018*, 1007–1014. https://doi.org/10.1145/3184558.3191531

Zhou, Y., Dredze, M., Broniatowski, D. A., & Adler, W. D. (2018). *Gab: The Alt-Right Social Media Platform*.

Primary sources

@a. (n.d.). Retrieved 18 April 2019, from https://gab.com/a

Community Guidelines. (n.d.). Retrieved 10 June 2019, from http://web.archive.org/web/*/https://gab.ai/about/guidelines

@gab. (2017, August 18). Retrieved 1 May 2019, from https://web.archive.org/web/20170818010419/https://gab.ai/gab/posts/9584661

@gab. (2018). Retrieved 26 May 2019, from https://gab.com/gab/posts/eWdsUTdKMDVDN2lCdXFIWVhTWElOUT09

Gab. (n.d.). Retrieved 5 February 2019, from https://gab.com/

@gab. (n.d.). Retrieved 5 February 2019, from https://gab.com/gab

Gab Help. (2018a, November 15). Creating, Viewing & Engaging With Content—Gab FAQ And Help. Retrieved 1 May 2019, from http://web.archive.org/web/20181115183638/https://help.gab.com/article/18-making-viewing-content

Gab Help. (2018b, November 15). Feed Filters—Gab FAQ And Help. Retrieved 1 May 2019, from http://web.archive.org/web/20181115183749/https://help.gab.com/article/29-feed-filters

Gab Help. (2018c, November 17). Gab Basics—Gab FAQ And Help. Retrieved 1 May 2019, from http://web.archive.org/web/20181117183610/https://help.gab.com/article/10-gab-basics

@getongab. (2018, October 28). Retrieved 30 May 2019, from The Wayback Machine website: https://web.archive.org/web/20181028015946/https:/twitter.com/getongab/status/1056362626077220865

List Of Banned Subreddits. (n.d.). Retrieved 13 February 2019, from https://www.reddit.com/r/ListOfSubreddits/wiki/banned

The Twitter Rules. (n.d.). Retrieved 10 June 2019, from http://web.archive.org/web/*/https://support.twitter.com/articles/18311

Twitter Blog. (2008a). News and Updates. Retrieved 26 April 2019, from https://blog.twitter.com/en_us/a/2007/news-and-updates.html

Twitter Blog. (2008b). Turning Up The Heat On Spam. Retrieved 26 April 2019, from https://blog.twitter.com/en_us/a/2008/turning-up-the-heat-on-spam.html

Twitter Blog. (2014). Another way to edit your Twitter experience: With mute. Retrieved 26 April 2019, from https://blog.twitter.com/en_us/a/2014/another-way-to-edit-your-twitter-experience-with-mute.html

Twitter Blog. (2019). A healthier Twitter: Progress and more to do. Retrieved 30 April 2019, from https://blog.twitter.com/en_us/topics/company/2019/health-update.html

Twitter Help Center. (n.d.). How to use advanced muting options. Retrieved 1 May 2019, from https://help.twitter.com/en/using-twitter/advanced-twitter-mute-options