

**PREVALENCE OF BIG TECH CENSORSHIP**

**IS CENSORSHIP RIGHT?**

A Thesis Prospectus

In STS 4500

Presented to

The Faculty of the

School of Engineering and Applied Science

University of Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science in Computer Science

By

Kiran Manicka

October 27, 2022

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

**ADVISORS**

Catherine Baritaud, Department of Engineering and Society

Madhur Behl, Department of Computer Science

In the past 5 years, the discussion of whether to censor hate speech has been one of big concern. On one end, many uphold that the right to free speech is one of utmost importance that is reinforced by the United States first amendment. In many cases historically, the first amendment's requirement of free speech has aided this country's marginalized groups (Wikipedia Publishment, 2022). People who would normally not have the ability to speak out could voice their concerns against more powerful groups. The first amendment's right to free speech rested in the heart of the civil rights movements as well as the United States relinquishment of Vietnam during the Vietnam war. There are many other smaller issues that this has helped with as well. Recently however, there has been an increase in the use of this right to propagate all sorts of hate speech. One of the most famous occurrences of this is how Alex Jones has used his platform to accuse the parents of the Sandy Hook shooting to be actors. These accusations have led to adverse effects for these families which include harassment and death threats from Alex Jones fans. Another example of this is the riots and insurgence of the Capitol Building in Washington D.C. which was fueled by Donald Trump when he refused to concede that he lost the election on fair terms. There are many more examples of this that have led to negative impacts for one or more marginalized groups or have simply just spread false information which is never beneficial for any society. Many people in recent years have recognized the importance of the first amendment, but have also juxtaposed that to the harmful effects of hate speech. As a result there has been controversy as to how to handle this law. This issue is delicate because the first amendment is something that is held so closely to every American citizen's heart. After all, it is the very first amendment that was listed by the founding fathers: the strongest foundation of this country.

The rise in technology as well the emergence of the internet in the previous decades has made this issue much more controversial. There are many popular content creators that have been dubbed the term influencers because of the vast amount of sway they have over their audience. The internet and social media have allowed influencers to propagate their individualistic ideas on a level never seen before. 30 years ago, if someone had a certain radical opinion, it would have never been spread as fast as it can be now. This amount of influence from individuals has imposed a very tough position for the tech industry. Tech companies like Twitter and Meta who run social media platforms are mostly part private institutions. Private Institutions refers to the idea they are run separately from the government and not in the sense that they are owned privately. When false news and hate crimes occur from their platform, society scrutinizes their businesses and puts part of the blame on them for allowing these messages to exist. Twitter and Meta have responded to this by taking a strict anti-hate speech policy towards all users on their platform. In the previous few years, these platforms have been very strict with hate speech and have even banned many influencers completely from their platform (Wagner K., 2021). However, the problem with this is that allowing companies to do this, allows commercial priorities to be inserted into the everyday lives of people (Cobbe J., 2020). This has led to even more controversy as to whether this is the right course of action to take and whether this is ethical. This controversy will be the foundation of the research laid ahead in this paper. There are two deep questions that can be derived from this that are tied together by this dispute. The first is whether hate speech on these tech platforms actually plays a significant and actionable role in harming people. The answer to whether certain hate crimes and false ideas in the United States is

a real noticeable trend from the hate speech on tech platforms, or whether they are one off instances that don't have any correlation to large scale trends. The second question is a lot more subjective and something that is a lot more deep set than the first: Is Censorship in the tech industry ethically correct? With this latter question being so subjective, the research behind this will largely be conducted off of trending opinions of society and as well as those from big figures in society like Tech Leads and Politicians. However these opinions will be coupled together with research to find an answer with a technical basis.

## **PREVALENCE OF BIG TECH CENSORSHIP**

Tech Companies taking the action to censor and ban certain influencers on their platform is only a justifiable course of action if there are clear correlations between certain speech and actionable events outside of the platform. Some might argue with this statement, however, if culpable speech on these platforms leads to no real correlation between hateful action, then the actions of Twitter, Meta and other tech companies to censor others is a clear violation of the free speech clause in the first amendment. The significance of this question is that, to find the objective answer behind this could serve as guidance for these Tech Companies. From a business standpoint, it is understandable why Tech companies would want to be strict with "hate speech". This is because these companies care about their reputation. Majority of the consumers of these platforms are from a younger generation that is considered more progressive. Therefore, these companies would want to cater to these people by taking a strict policy against certain types of rhetoric. This in turn would lead to more satisfied consumers and more revenue for the business. The one thing to note however, is that companies behaving this way are still infringing on first amendment rights. An institution infringing on the rights of another to benefit financially is

definitely a course of action that is liable to prosecution since that in of itself is a violation of law. Some might say that since these tech companies have a right to censor whoever they want to protect their revenue stream and wellbeing as an enterprise. While this is true, a lot of platforms like Twitter and Facebook have built such an immense and passionate following to the point where they are considered the status quo of media. This means that even if people did not agree with the way they conducted business and wrote internal policy, people would still continue using their products since they are the social norm now and the only way to stay “connected”. As can be seen from the statements before, finding the answer on whether certain speech truly does have an impact on the outside world on these platforms is integral to answering the overall question of whether Tech Companies should censor others.

The approach that I will be taking to find the answer to this is mainly by performing online research. I will aggregate multiple online resources regarding so-called hate speech on social media, as well as their outcomes to find an objective correlation to this. Many could call this a very weak and non-technical approach of research since I will be compiling the opinions of others, but I feel this is the only way to truly find a technical and objective correlation in the subjective feelings of the masses. I will also be analyzing certain headlines of popular influencers like Alex Jones and Andrew Tates to find key ties to certain actions taking place (Sung, 2022). Another resource that I will utilize to find myself a conclusion is census statistics of hate crimes. An example of this will be looking at the top ten most recent hate crimes to take place in the United States, and to do research into each one of their causes and motivations. For example, the riots on the Capitol Building is something that I would consider as an actionable event that had clear roots in fake news and social media rhetoric. On the other side of the coin,

the Sandy Hook shooting is an example of a person who acted in their own interest without significant influence from any social media platforms. Some might say that a lot of hate crimes cannot be traced back to a single point of cause so easily but could be the result of a small amount of influence from certain creators over a long period of time. While this might be true, this does not make it justifiable then to censor those creators and infringe on their first amendment rights.

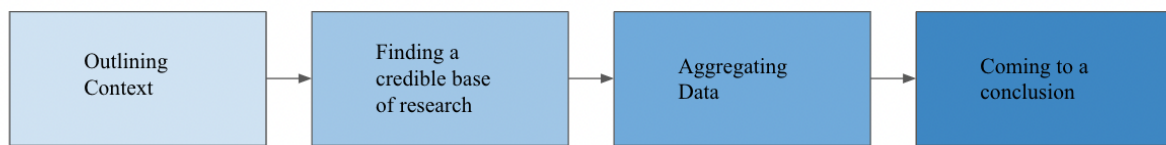


Figure 1: Tech Platform to actionable events research process. This figure visualizes the expected research process to find a correlation between general speech occurring on Tech platform and actual events that cause harm for the technical project.

The flow of the research will go as such; the first step will be to outline the context. The main objective of outlining the context is to set criteria for research instances. The reasons for this is that allowing hate crimes that did not have motivation from tech platforms would be superfluous to my research. The main criteria is that the hate crime/action needs to be a direct and unambiguous result of a post online on any platform. The second step of the research is finding a credible base. This helps ensure that subjectivity is minimized since there are any online opinions that float around. An example of this would be to use FBI censure statistics to determine the most recent hate crimes that have occurred. The next step is to aggregate data. This means to take the few dozen pieces of data that I have acquired from step 2 and to pool all of it in

one place. The last step would be to detect some sort of correlation and come to an objective conclusion.

The intended outcome is to be able to aggregate as much data to create some sort of conclusion. I hope to be able to find a strong enough correlation or a sufficiently scattered correlation to either justify or disagree with Tech Media censorship. The type of paper that I will write will be a technical research finding. This type of paper will also resemble a scholarly article. The hope is to eliminate all subjectivity from this matter since this research topic does have an opinionated nature. I also hope to motivate others to think about the research and what this problem means to them.

## **IS CENSORSHIP RIGHT?**

The second topic that will be covered through my research is whether censorship benefits or harms a society and how Big Tech plays a role in this. Naturally this topic will be more subjective than the previously mentioned topic since it relates to how people view politics, but just like my Technical Research, I will try to approach the question from a technical point of view. The question of whether censorship is right is really important since it dictates entire societies on how to act. Historically, the over limitation of speech in society has led to fascist and malicious authorities of power. However, the under limitation of speech has also had a disastrous impact and has allowed for certain views of prejudice and malice to be spread. This seesaw of ethics is what makes this question so difficult but so important for any society to address since it can dictate where a nation will head in coming decades. I believe this problem needs more research since there are so many that are conflicted between the two sides. I would like to present

my own technical understanding of the question at hand and convey my findings through objective facts as well as historical basis. The approach that I will be taking is to analyze historical references and examples as well as the modern ones to see the impact that limiting free speech has. I will utilize the Actor Network Theory(Fenwick et al., 2010), since the act of censorship starts off in a small localized network, but its impacts can spread into larger networks. An example of this how censorship start off small in 1940's Nazi German, but then grew to be a large movement which led to fascist government.

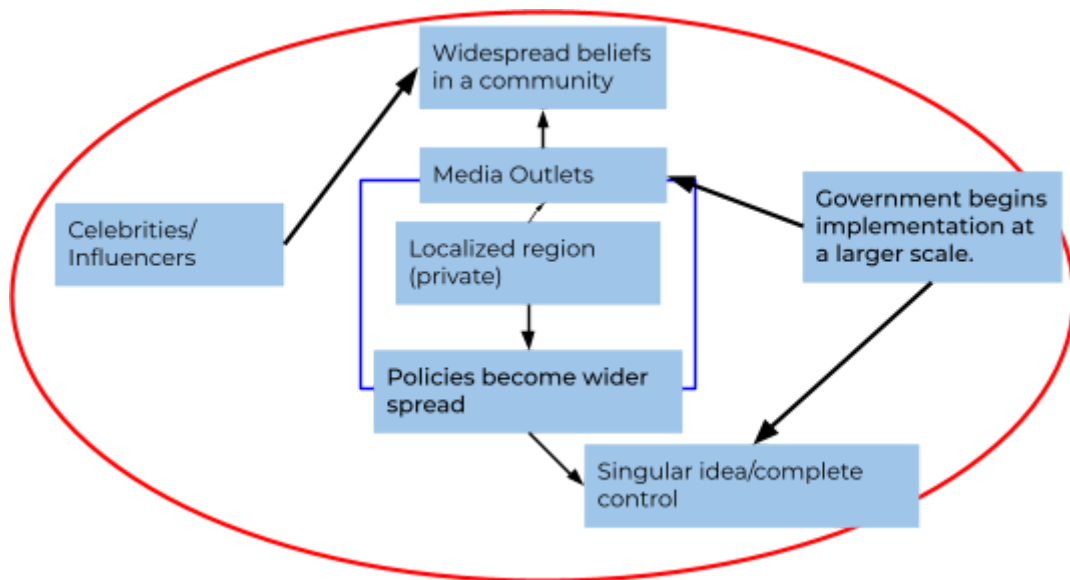


Figure 2: Hate Speech Propagation, ANT. This figure visualizes the theoretical expansion of hate speech that starts on tech platforms which will be tested in the technical and STS project.

In Figure 2, multiple entities and courses of action are listed. As can be seen, institutions like the government and the media can take localized instances of censorship and turn them into widespread beliefs and policies that are enforced. What I want to analyze as part of my STS



research is which of these connections (arrows) in Figure 2 are tenuous relationships and break the theorized ANT listed.

The outcome that I hope for is similar to the previous topic's. I want to be able to find either a strong correlation between small level censorship(similar to how removing a few Tweets or banning a user on Facebook is considered relatively small) and large-scale negative impacts on society or a weak scattered relation between the data. The conclusion that I find, could help guide how Big Tech can play a role when dictating their policy. The type of paper that I will write will be a technical presentation of findings with a slightly subjective conclusion that summarizes the topics of that paper. I would also include graphics and visuals that can present how the Actor Network Theory plays a role in the growing impacts of censorship if there is one.

### **INTENDED IMPACT?**

The debate over free speech and the Tech Industry's role in limiting it has been one of the most contentious discussions in recent years. The action of one industry could influence the direction of an entire nation just as it did in previous civilizations. Truly understanding this subject area is significant. To really grasp the context and facts about this discussion, one must be informed of whether hate speech on tech platforms actually leads to actionable hate crimes and if censorship benefits or harms a society. That is the aim of my research. Without a solid understanding, simple words on a screen can turn into chaos, and the foundations that make up a nation can crumble.

## REFERENCES

- Cobbe, J. (2020). *Algorithmic Censorship by Social Platforms: Power and Resistance*. Philos. Technol. <https://doi.org/10.1007/s13347-020-00429-0>
- Fenwick, T., Edwards, R. (2010). *Actor-Network Theory in Education*. Routledge. <https://doi.org/10.4324/9780203849088>
- PubAffairs Bruxelles. (2021, December 9). *The commission proposes to extend the list of 'Eu crimes' to hate speech and hate crime*. PubaffairsBruxelles. <https://www.pubaffairsbruxelles.eu/eu-institution-news/the-commission-proposes-to-extend-the-list-of-eu-crimes-to-hate-speech-and-hate-crime/>
- Rubin, J. (2020, July 2). *Tech companies are finally being shamed into action*. American Academy of Arts & Sciences. <https://www.amacad.org/news/tech-companies-are-finally-being-shamed-action>
- Sung, M. (2022, August 22). *Andrew Tate banned from YouTube, Tiktok, Facebook and Instagram*. NBCNews. <https://www.nbcnews.com/pop-culture/viral/andrew-tate-facebook-instagram-ban-meta-r-cna43998>
- Surfsharkadmin. (2022, April 12). *Social media censorship cases decreased by 35% in 2021*. Surfshark. <https://surfshark.com/blog/social-media-censorship>

Wagner, K. (2021, July 14). *Twitter penalizes record number of accounts for hate speech*. Time.

<https://time.com/6080324/twitter-hate-speech-penalties/>

Wikimedia Foundation. (2022, September 1). *Hate speech in the United States*. Wikipedia.

[https://en.wikipedia.org/wiki/Hate\\_speech\\_in\\_the\\_United\\_States#:~:text=In%20a%20Supreme%20Court%20case,basis%20of%20the%20speaker's%20viewpoint.](https://en.wikipedia.org/wiki/Hate_speech_in_the_United_States#:~:text=In%20a%20Supreme%20Court%20case,basis%20of%20the%20speaker's%20viewpoint.)