

Smart Charlottesville – Changing our Community Interactions

(Technical Paper)

The Effects of Facebook's Response to Content Moderators

(STS Paper)

A Thesis Prospectus Submitted to the

Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree
Bachelor of Science, School of Engineering

Cory Ayers
Fall 2019

Technical Project Team Members

Anthony Lancaster

Conner Hutson

Jared Tufts

Kajal Sheth

Luke Deni

Sanjana Hajela

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

Signature Cory Ayers Date 11/1/2019

Cory Ayers

Approved Ahmed Ibrahim Date 11/24/19
Ahmed Ibrahim, Department of Computer Science

Approved Kent Wayland Date 12/12/19
Kent Wayland, Department of Engineering and Society

General Research Problem

How can websites promote positive online interactions while protecting against explicit content?

Social media and the internet have greatly changed the way we communicate with others in the 21st century. People can now hide behind screen names and talk to others with little to no consequences. Over time, the interactions we have online have uncovered many new issues, including cyberbullying, seeing graphic images/videos, and hate speech being spread. These interactions can have long-lasting effects on the people who witness them and it is a problem that needs more attention. Every year, millions of posts online are removed due to having graphic or inappropriate content (Rosen, 2018, p. 1). This does not mean that online interactions cannot have positive effects on humans and their communities. When fostered correctly, the internet can be a place that brings people together and promotes positive experiences for users. Collaboration with others through the internet can allow companies to do business with each other across the world and other feats that were impossible before computers. This however, can only result if the sites being built are carefully created and monitored to promote healthy interactions. I will be examining the effects of harmful posts online, and why combating these posts is an extremely important issue in protecting not only the users of the site, but the humans who help moderate this negative web traffic.

Technical Research Problem: Smart Charlottesville - Changing our Community Interactions

How can our team create a website that brings community members together to advance the future of Charlottesville in a positive way?

The implementation of technical online platforms has become an increasingly popular idea to engage residents of a city with local government, and the University of Virginia plays a vital role in this due to its technical expertise. Professor Ferguson and Professor Ku are part of the STS department at the University of Virginia, and they are conducting research with their STS 4500 students to develop research ideas for transforming Charlottesville into a smart city. These ideas, however, need a platform that can be viewed and contributed to by both residents and the local government. The research problem to be solved is how to efficiently communicate these ideas, and others, to the Charlottesville community to improve the city for the future.

Currently, there is no viable platform that solves this problem of lacking communication in Charlottesville, since the research problem demands different user types and custom databases that are unavailable with platforms such as WordPress. The work done by the current capstone group of this academic year will provide the first iteration of a solution to bridge the gap for collaboration between the university and Charlottesville. The capstone project will last the entire academic year of 2019-2020.

Creating a web application public to all users will address the problem in a positive way, by working to provide a safe, non-anonymous site for community members to share ideas of changes they want in the community through engaging discussions. The website will have a feature for users to submit blueprints for proposed projects where they can also add file attachments such as pictures. The platform will require users to register and login to submit posts and interact with other users. The users will be able to look at the projects and comment on them, mark projects as “favorites” for easy access later, and connect with the authors of the blueprints via email. Community members can also submit smaller problems around the city to gain attention from other members so they can be fixed. There will be an “about us” tab where interested visitors can get in contact with the creators of the site and learn more about this initiative. The landing page will have a map that shows the Charlottesville area with ongoing projects pinned so users can explore projects in different areas by clicking specific pins on the map. Finally, there will be a resources tab describing places users can go to learn more about projects and current city work in general.

To build this website a strict set of requirements will be collected from Professor Ferguson and Professor Ku. Requirements contain the attributes and properties of features of a system that the user wants to help solve their problems. It is important to gather system requirements to correctly understand the goals of the client and to facilitate the work of the developers to best cater towards the stakeholder’s needs. Listed below are the capstone group’s minimum, desired, and optional requirements:

Minimum Requirements:

1. As a user, I want to be able to comment on a blueprint to give my support or feedback.
2. As a user, I want to be able to filter through blueprints based on what category they fall under.
3. As an administrator, I should be able to manage blueprint content by hiding or removing it.
4. As an administrator, I should be able to manage the privileges of other users (students, community partners, and community members).
5. As a student, I should be able to create my own blueprint space so that others may view it.
6. As a student, I should be able to view other student’s blueprints.
7. As a community member, I should be able to leave comments on a student’s blueprint.
8. As a community member, I need to be able to post blueprints.
9. As a community member, I need to be able to like specific comments or blueprints.

Desired Requirements:

1. As a user, I should be able to search for keywords that define the type of blueprints posting I want to look at.

2. As a user, I should be able to view blueprints based on specific location

Optional Requirements:

1. As a user, I should be able to comment on other comments.
2. As a student, I should be able to tag my post with specific categories.

At the end of the project, we will have a collaborative, online workspace reachable by both the Charlottesville community and academics at the university. Users will be able to post ideas, gather feedback, collaborate, and connect with university resources; the university can do the same, as well as be able to identify problems in the community that may have otherwise remained hidden.

STS Research Problem: The Effects of Facebook's Response to Content Moderators

How has Facebook's response to public backlash over content moderators changed the lives of their workers, and what results have these changes produced?

Introduction:

The interactions we have on social media every day are not guaranteed to be positive ones. Cyberbullying, graphic sexual images, and threats can all happen and may have lasting effects on different people. This can especially be true for the content moderators of Facebook, one of the world's largest social media companies. The workers who look through explicit and damaging content for endless hours each day may be the most susceptible to having their long-term mental health affected. Recently, information has come out about Facebook's role in creating these hostile work environments and how they have done little to change issues they knew were present (Newton, 2019, p. 5). Content moderators complain they do not receive enough psychological services, and that they are not allowed to talk to friends due to strict confidentiality agreements (Dwoskin et al., 2019, p. 2). The public has gained information through former employees breaking non-disclosure agreements (NDAs), but the secrecy surrounding the subject and different offices around the world means there may be much more going on (Newton, 2019, p. 4). This knowledge gap could have extreme costs, especially if Facebook attempted to hide what is still unknown to the public. Stronger NDAs and other legal requirements could give future employees even less rights, and make them susceptible to harmful legal action if they break them. Moving forward, I want to examine Facebook's response and what they are currently doing to respond to the backlash.

Background and Theoretical Framework:

The overall health of Facebook moderators has clearly been shown to be affected by this job through countless employees speaking about their experiences. For example, in Casey

Newton's examination of Facebook moderators he looks into one man's experience stating, "The job also changed the way he saw the world. After he saw so many videos saying that 9/11 was not a terrorist attack, he came to believe them." (Newton, 2019, p. 33) Other workers now claim to sleep every night with a gun in their hands, one even sweeps his house every morning in fear that someone may be inside (Newton, 2019, p. 33). These instances are not only occurring in the United States, but worldwide at different offices. In Asia, one employee discusses how skyscrapers give him constant flashbacks to all of the suicides he has reviewed (Dwoskin et al., 2019, p. 18). Even though he has quit his job, constant therapy and special assistance is still needed to help him get through everyday life (Dwoskin et al., 2019, p. 18).

These results are clearly serious and significant, as seen through studies done by UCF and Penn State professors. They have examined how negative online experiences can lead to symptoms of Post-traumatic stress disorder (PTSD) in victims, including looking at explicit/sexual content and cyberbullying (McHugh et al., 2018, p. 1). Having these interactions every day might take an even worse toll on workers, as shown in these examples.

It is extremely important to examine these stories and look at Facebook's benefits and policies surrounding this situation. Many former employees stated that they were staring at content non-stop and that they were strongly encouraged to examine a certain amount of content each day (Newton, 2019, p. 16). Other employees complain of a lack of counseling available and that they are limited to very few breaks during the day (Newton, 2019, p. 23). Part of this is a response to thresholds set by Facebook, that ask for moderators to judge posts correctly a stunning 95 percent of the time (Newton, 2019, p. 14). Managers at these offices also examine the amount of content each employee goes through, and their jobs could be at stake if they do not examine content fast enough (Newton, 2019, p. 16). This adds more stress to these intense jobs.

While Facebook helps set these policies, the workers are not actually in Facebook offices. Facebook, as well as other tech giants, use contracting companies, such as Accenture and Cognizant, to hire these employees to further protect themselves from the liabilities of content moderator health (Dwoskin et al., 2019, p. 6). In the past few years, plenty of information has come out to the public regarding this issue, largely due to former employees breaking NDA agreements. These former employees still felt the long-term effects of this psychologically straining job and felt it was their duty to speak up to protect future employees. I will be examining the responses of Facebook, contracting companies (such as Cognizant and Accenture), and the actual workers when looking at this case, as well as the public response to these changes if relevant. When examining the sociotechnical system here, the overall goal is an equal emphasis on the technical aspect of Facebook as well as the human aspect. I want to look at this balance more and see how Facebook is responding to the criticisms of imbalance in this system and how technological advancements (such as artificial intelligence) might be hindering their hopes of striking this balance. The response by Facebook will directly impact the contracting companies they use, especially if they impose new standards for employee health. In the past week, Cognizant announced it would be terminating its current work with Facebook due

to concerns over moderators (Carbone, 2019, p. 1). If they follow through with this, it could drastically change Facebook's future actions and how they look to respond in a positive way.

Data Analysis:

To address this compelling question, I will need to gather evidence and data involving not only the responses by Facebook and contracting companies, but around the previous situation to compare how they have changed. Within my background, I have already done research on the work lives at different moderating offices, but I want to look deeper into the response. Just last year, Facebook started releasing information on content moderation and it will be interesting to see if they focus more on the content or the moderators (Lapowsky, 2019, p. 2). I plan to find and analyze this data to see how content moderation statistics are changing, as well as Facebook's comments on the results these are having on employees (Rosen, 2018, p. 1). I also will be looking to collect more accounts of testimonies from current employees that are still speaking out and discussing what has changed over the past several years. This task may prove difficult, as this involves them breaking legal agreements, but it has happened plenty in the past. These are the different ways I plan to learn about this case, to effectively produce a case study with solid evidence based on current facts regarding the issue.

When looking at these different testimonies of current and former employees, I will look to analyze similar factors between the two and how they have changed since Facebook acknowledged the content moderator issues. I plan to also look at the statistics Facebook has been releasing, and see if changes over the years could be the results of their response to the backlash. For example, will they be posting figures only about content moderation effectiveness rates going up, or will they directly have reports on the wellness of these employees. This will be interesting to evaluate because effectiveness rates going up does not necessarily mean they are treating moderators better. Facebook and the contracting companies they employ seem to be looking to take steps to change the current roles of moderators, and I look forward to seeing what these are going forward.

Summary of Project:

By the end of my project, I hope to be able to show how Facebook's response to content moderator backlash has changed workplace practices for its contractors and whether these changes are making their jobs easier or more difficult. By then I will have more resources on the data from Facebook as well as more testimonies from people with direct knowledge of the situation. This information will be extremely useful for content moderators living in different conditions around the globe, but also for those concerned about the lives that are currently being affected by this occupation. If the results from this study show Facebook is making changes that they believe will help humanize this job and improve the lives of their moderators, knowing these could be useful for changing the culture at other large technology companies.

Conclusion

This work will hopefully be able to create more awareness on the topic of interactions we have online and how significant they can be. Even the smallest interactions and content you put online can potentially affect others in extreme ways. People rarely think about the workers who have to look at explicit content and the effects posts can have on their mental health or those of other users. The importance of this cannot be overlooked, as it involves hundreds of thousands of people around the world. People could continue to look at my research to study the progress of large media companies and their work in improving the lives of their users and workers. By examining not only the user experience, but also the employee experience, companies may learn far more about problems they should be looking at. All of this is an effort to improve social interactions online and the results that can occur from negative interactions.

Bibliography

- Cabato, R., Dwoskin, E., & Whalen, J. (2019, July 25). Content moderators at YouTube, Facebook and Twitter see the worst of the web - and suffer silently. Retrieved from <https://www.washingtonpost.com/technology/2019/07/25/social-media-companies-are-outsourcing-their-dirty-work-philippines-generation-workers-is-paying-price/>.
- Carbone, C. (2019, November 1). 6000 jobs go as Facebook loses filter. Retrieved from <https://www.news.com.au/technology/online/social/company-that-moderates-facebook-exits-amid-concerns-over-mental-health-of-moderators/news-story/3eb3c862852d75ff0b6d94abd251783a>.
- Carroll, J. M., McHugh, B. C., Rosson, M. B. & Wisniewski, P. (2018, October 2). When social media traumatizes teens: The roles of online risk exposure, coping, and post-traumatic stress. *Internet Research*, 28(5), 1169–1188. <https://doi.org/10.1108/IntR-02-2017-0077>
- Lapowsky, I. (2019, May 23). More Data on Content Moderation Won't Silence Facebook's Critics. Retrieved from <https://www.wired.com/story/facebook-community-standards-report/>.
- Newton, C. (2019, February 25). The secret lives of Facebook moderators in America. Retrieved from The Verge website: <https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona>
- Newton, C. (2019, June 19). Three Facebook moderators break their NDAs to expose a company in crisis. Retrieved from <https://www.theverge.com/2019/6/19/18681845/facebook-moderator-interviews-video-trauma-ptsd-cognizant-tampa>.
- Rosen, G. (2018, May 15). Facebook Publishes Enforcement Numbers for the First Time. Retrieved from <https://newsroom.fb.com/news/2018/05/enforcement-numbers/>.