Debugging and Implementing new Features in a Production Codebase
(Technical Paper)

The Contribution of Machine Learning Algorithms to Radicalization
(STS Paper)

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
Andre Knocklein

Oct 8, 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

Joshua Earle, Department of Engineering and Society

Rosanne Vrugtman, Department of Computer Science

Introduction

For my technical report, I will be discussing my experience working as a Software Engineering Intern at SitScape in the summer of 2022. This report will have a focus on implementing new features and debugging old ones in a production code base meaning the work I did was viewed by the customer. It will focus on my perspective as an entry level developer who did not even have experience with the programming languages that were used and how that lack of experience influenced my experience. It will also cover the techniques I used to overcome my lack of knowledge and how that might be applicable to other entry-level software engineers.

For my STS report, I plan to discuss the extent to which machine learning algorithms used by content-serving companies influence radicalization in their users. Every year, society becomes more and more hooked on internet content and social media (Pew Research Center, 2022). This gives the companies that serve that content immense power to influence their users. When combined with the goal of maximizing the time users spend on these websites and the introduction of machine learning algorithms that are so secretive and complex that there is most likely not a single person that completely understands them, the radicalization of some groups of users on these platforms has been observed. To discuss this radicalization is more important now than ever because the political landscape, especially in the United States, is becoming more and more polarized, and discussing a potential cause is important. The most common form of this radicalization has been dubbed the ‘Alt Right Pipeline’ and this report will discuss how prevalent it is, what the entry points to it are, and what groups are most likely to fall into it.

Technical Project

Abstract:

A software company based in Northern Virginia had an expansive codebase in continuous development with the desire to fix customer-reported bugs and implement customer-requested features. The variety of tasks I was assigned to in a variety of languages using a variety of tools allowed for the opportunity to use a general debugging methodology. This involved the use of tools I had not used before like a Linux machine to host the local build and Google's site debugging tools. Using these tools allowed for an effective way to understand a codebase far too large to comprehend by traditional reading of the code. The methodologyI used to understand the codebase allowed for the implementation of a dozen bug fixes and a few new features that are currently part of the product. The codebase still has bugs, and there will be a continuous maintenance of the code, as is the nature of production code.

STS Project

Research Question

The technology at the center of this STS report is machine learning. Machine learning is a growing field in computer science which is concerned with using data to improve performance on a given task. Specifically, I will examine the usage of machine learning in creating algorithms for content-serving websites. These websites use these algorithms to recommend content that is thought to be relevant to the given user by using the data of what that user has consumed previously and what similar users have consumed in order to recommend new content or tailor a search query to that user. Since these algorithms use data, in many cases without supervision from humans, in order to improve themselves, the actual nature of how they work or what they

will produce is hidden. For example, YouTube uses a Deep Neural Network which is a machine learning algorithm which is a complex algorithm which, at the given scale of data that YouTube uses, cannot be supervised (Covington, 2016). The term supervised in machine learning means that there is a human guiding what the algorithm does, in this case, there is no feasible way for a person to guide the algorithm because the amount of data it combes through is much too large. Therefore, this algorithm runs by itself, and people have some high-level understanding of how it works, but the exact reasoning as to what it does is either not known or difficult to figure out. Combining this with the fact that companies are very secretive about the algorithms they use makes the internal working of the algorithms impossible to know, but the consequences they create can be readily observed.

My research question is to what extent do these algorithms contribute to the radicalization in political ideology of the users of the websites that use them. I will discuss the nature of this radicalization as well as how it typically gets started for users. I will also discuss how this radicalization has been coined the “Alt Right Pipeline”. With an ever-increasing portion of the population using the websites that use these algorithms and an ever-increasing sense of polarization (Geiger, 2021) and radical rhetoric, it is important to research one of the ways in which this polarization gets started in those websites. Beyond just the knowledge of how this radicalization works, it is important to realize its harmful effects on the groups of people targeted by radical thought.

Relevant Social Groups

Radicalization in all its forms targets a group of people that is painted as the enemy, even when there is no evidence to show that that group represents a legitimate grievance against the

radical group. With the radicalization through these algorithms, the groups that are targeted most commonly are the groups that are already marginalized in society, so basically any group other than white males. The rhetoric that is espoused by the content that these algorithms serve up serves to paint marginalized groups as the cause of some perceived problem to the group that is consuming the content. This kind of rhetoric is very close to hate speech which leads to political radicalization (Bilewicz, 2020). Furthermore, it causes the consumers to use similar rhetoric which has direct negative consequences to the people belonging to marginalized groups in contact with these consumers, but it also has a wider effect on the overton window of what is socially acceptable rhetoric towards marginalized groups. This can increase the hateful rhetoric towards marginalized groups even from those that didn't consume this content.

The other side of the coin is that the people being radicalized are also negatively affected by the effects of the algorithm. These, mostly young white males, are being driven towards hateful ideologies just based on the data an algorithm has and what it thinks they might like. There is some blame to be placed on this group, but I would argue that they are also victims of this effect which they cannot completely control. The ideologies that they are being pushed towards will never actually help to improve their lives.

It is important to consider both sets of groups in order to gain a full understanding of the effects of this because only then can a complete solution be proposed. One set of groups is clearly more of a victim in this relationship, but it is important to think of the other group too because that is where the most insight into the problem will be gained.

Another group that should be considered is that of the content creators that are also responsible for the radicalization. Content creators are just as much affected by the pressures of the algorithm as the users. Content creators will create what gets them the most clicks or views,

and often if they can find a niche, even if it is harmful, they can reliably stay there and make a living because the algorithm will continue to recommend their content to a group that will reliably consume it. The nature of the content that is created by this group is important to analyze as well.

Frameworks and Methods

In order to analyze the rhetoric that is prevalent in the problematic content that is recommended by the algorithm, I will use a discourse analysis framework. This will provide insights into what exactly the rhetoric is that is promoted by the algorithms, and why that rhetoric is appealing to the group of users that consume it. It will also be useful in understanding how the content influences the nature of discourse outside of online spaces of the topics that are prevalent in the recommended content.

I will also use a case study framework in order to illustrate an example of how radicalization normally unfolds. This will allow me to detail each step in the process of radicalization and discuss how the machine learning algorithm plays a part in allowing that step to happen. It will also allow for specific examples of pieces of content in which the rhetoric is problematic at each step of the process. Ultimately, it will provide a greater understanding of the problem as a high level view of how it works becomes clearer.

While using the case study framework, I will use a “The Stories We Tell Ourselves” methodology. I will analyze the content that the algorithm recommends to find out what the content creators are really trying to say with that content. I will analyze who that content proposes is the hero of stories it tells, and more importantly, I will analyze who that content paints as the villain. This is where the real danger of the content becomes clear as this seems to

often be what the consumers of the content latch on to. The villains of the stories in the content will often become the enemy of the group of people that consumes that content.

I will expand this methodology in order to uncover the overarching story that the algorithm itself seems to tell to its users on the path to radicalization. This will be an aggregate of the individual stories that the promoted content and the values that they espouse. This aggregate story will be the model that most likely shapes wider discourse and will answer the question of how much the algorithms actually influence radicalizations.

Key Texts

One of the most useful resources for discussions of this kind related to how algorithms feed into systems of oppression and radicalization will be "Algorithms of Oppression"(Noble, 2018). It focuses on how Google's search algorithm feeds into systems of oppression that are already present in society. It is a useful resource for this project because it discusses how the algorithm and the lack of willingness to oversee it, can have consequences towards marginalized groups. My discussion in this project will be similar, but I will focus more on the rhetoric of users that consume the content than the actual content itself.

A powerful force in the radicalization of users on YouTube specifically is PragerU. This company uses a facade of an educational institution in order to serve its viewers right-wing propaganda. It has a well established presence on YouTube, and the article "The Kids Are Alt-Right: An Introduction to PragerU and Its Role in Radicalization in the United States" (Dickinson, 2021) will be a useful tool to analyze its impact. The manner in which PragerU influences its viewers and the rhetoric it uses is similar to other content creators that have a presence in the alt right pipeline, so it is a useful case study. PragerU also features many other,

similar content creators on its content which makes it a good measure of how users can fall into its rhetoric.

In order to dive deeper into the recommendation algorithm that YouTube uses as a case study, the paper “Deep Neural Networks for YouTube recommendations” written by Google engineers will be very helpful (Covington, 2016). It will provide the technical information needed to explain how these algorithms actually work on a high level which will be a valuable insight in discussing the effects of these recommendations.

Finally, it will be useful to know the strategies that radical groups use in order to indoctrinate new members. This is where the book *Indoctrination to hate: Recruitment techniques of hate groups and how to stop them* (Dunbar, 2022) will be useful. Knowing this information, a comparison can be drawn with the content that leads into the Alt-Right Pipeline in order to discover to what extent that content is a radicalization technique.

References

1. *Social Media Fact Sheet*. (2022, November 16). Pew Research Center: Internet, Science & Tech. https://www.pewresearch.org/internet/fact-sheet/social-media/?menuItem=3814afe3-3f3c-4623-910b-8a6a37885ab8
2. Covington, P., Adams, J., & Sargin, E. (2016). Deep Neural Networks for YouTube Recommendations. *Proceedings of the 10th ACM Conference on Recommender Systems*. https://doi.org/10.1145/2959100.2959190
3. Geiger, A. (2021, April 9). *Political Polarization in the American Public*. Pew Research Center - U.S. Politics & Policy. https://www.pewresearch.org/politics/2014/06/12/political-polarization-in-the-american-public/
4. Bilewicz, M., & Soral, W. (2020). Hate Speech Epidemic. The Dynamic Effects of Derogatory Language on Intergroup Relations and Political Radicalization. *Political Psychology*, *41*(S1), 3–33. https://doi.org/10.1111/pops.12670
5. Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. Amsterdam University Press.
6. Dickinson, R., & Cowin, T. (2021). The Kids Are Alt-Right: An Introduction to PragerU and Its Role in Radicalization in the United States. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4187075
7. Dunbar, E. W. (2022). *Indoctrination to Hate: Recruitment Techniques of Hate Groups and How to Stop Them*. ABC-CLIO.