

## **Thesis Project Portfolio**

**Debugging and Implementing new Features in a Production Codebase**

(Technical Report)

**The Contribution of Machine Learning Algorithms to Radicalization**

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

**Andre Knocklein**

Spring, 2023

Department of Computer Science

## **Table of Contents**

Executive Summary

Debugging and Implementing new Features in a Production Codebase

The Contribution of Machine Learning Algorithms to Radicalization

Prospectus

## **Executive Summary**

### **Introduction**

The two primary documents in this portfolio are a technical report and a thesis research paper. The topics of the two are not linked by more than the fact that they discuss issues relating to computer science. The technical report is about my experience at a software company as a novice programmer, and the thesis is about the role that machine learning algorithms play in radicalization pathways in online spaces.

### **Technical Report: Debugging and Implementing new Features in a Production Codebase**

The technical report in this portfolio is about the process of debugging and implementing new features in a production code base from the perspective of a novice programmer. This paper starts by introducing my experience with the topic from when I was working at this software company as an intern. It then discusses some of the problems that I faced like my lack of experience and the size of the codebase. After this, there is a literature review of papers relating to debugging.

After this introduction, I outline my process for the tasks that I needed to complete. These tasks are split into two sections: Debugging and New Features. The process for debugging is split into five steps and each one is detailed in the report. These steps are problem definition where the given bug is found and reported, cause identification where the problem code is pinpointed,

debugging which is the step where the fix is created, result verification where the new changes are checked, and quality assurance process in which the final checks occur in terms of the rest of the code base and where the changes are merged into the codebase. The process for creating new features is similar as the final two steps are the same as in debugging. In the new feature creation process the first step is requirements elicitation where the specifications of what the feature exactly should be are developed. The next step is the implementation process in which the code is written to build the feature.

The report next touches on the outcomes of the tasks I did which were varied, but the quality of life improvements were the most useful changes. Then it discusses what I learned from the experience which consists mostly of mastery of JavaScript and a good understanding of the jQuery library as well as a bit of knowledge about the PHP programming language. Finally, it describes how I will continue working on similar tasks in the future and how this experience is a useful starting point.

### **Thesis: The Contribution of Machine Learning Algorithms to Radicalization**

The thesis begins with an introduction to both the type of extremism that will be the focus of this paper and the primary phenomenon that this paper covers. The type of extremism is termed as the “alt-right” which is a right-wing ideology that focuses on cultural issues and is active in online spaces. The phenomenon that this paper focuses on is termed the “alt-right pipeline” and it is a system of increasing radicalization for users in online communities which has grown over the last decade mostly. Specifically, this paper is concerned with how the

creation of content-serving machine learning algorithms have impacted this pipeline and increased its severity. The primary research question of the thesis is to what extent do these algorithms contribute to the radicalization in political ideology of the users of the websites that use them.

The thesis then goes into a platform analysis in which the chronological development of the alt-right pipeline on different platforms is uncovered. The method of inquiry used to do this analysis is that of case studies into the platforms and content that contributes to the pipeline. The first appearance of alt-right material online is in bulletin board system websites. The primary focus of the thesis is on YouTube. It is the biggest content platform on the internet, and it is also where the trends of the alt-right pipeline are most apparent. On YouTube, there are three levels of increasingly extreme types of content which concerns this analysis. The paper analyzes how these levels lead into each other and how this trend causes users to be exposed to more and more extremist content. Examples of all three levels are taken into consideration in the analysis and the analysis concludes with a high level view of how the pipeline works on YouTube. The thesis also quickly touches on trends on TikTok which is where the pipeline will expand to next and already has to some extent.

The thesis then thoroughly analyzes the technology at the center of the recommendation systems that platforms use, which is deep neural networks, a type of machine learning. The example in this analysis is the YouTube recommendation algorithm. The analysis discovers that if one member of a demographic consumes some kind of content, the algorithm is likely to recommend it to new users because the algorithm takes demographic and viewing habits into account to group users.

The thesis combines the platform and technology analyses together to create a holistic picture of the problem. This picture is that content can be shaped in such a way that users gradually become radicalized in a way that is hidden to them, and once one user experiences this effect, more users are likely to have the same thing happen to them.

Finally, the thesis discusses the issue of whose responsibility this problem is and what can be done to fix it.