Evaluating detection mechanisms for misinformation spread on social media

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

Pablo Weber
Spring 2021

Technical Project Team Members
None

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  Date 05/14/2021
Pablo Weber

Approved _____ Date _____
Richard Jacques, Department of Engineering and Society

Approved  Date 11/08/2020
Aaron Bloomfield, Department of Computer Science

Over the past 16 years, the world has seen an increase in internet users from 413 million in 2000 to over 3.4 billion in 2016. Of those users, 34.6% are active on social media. Worryingly, in 2014, 61% of millennials in the US claimed to get their political news from Facebook, compared to just 44% claiming to get their political news from CNN (Moon). This reliance on social media for news has been taken advantage by actors looking to influence and manipulate people's thoughts with the usage of misinformation – false information deliberately spread to influence people's thoughts – colloquially referred to as fake news. The consequences of fake news have recently come to light with the creation of groups such as Qannon, an extremist far-right group who recently took part in storming the US Capitol. Clearly, fake news is an incredibly powerful tool and can quickly become devastating when used for nefarious reasons. It is for this reason that my portfolio focuses on curbing its spread.

The aim of my technical project is to alert social media users to a post potentially spreading misinformation. More specifically, I do this by creating a Chrome extension that overlays a widget on the top right corner of Tweets with an "accuracy score". Using Machine Learning, the Tweet is analyzed and cross-referenced with trustworthy news articles. The less similarities there are, the lower the score. It is my goal that this extension will not only warn but also inform users to potential fake news, lowering its negative effects on Twitter users.

My STS research paper takes a more general approach and identifies ways in which we can stop the actual spread of fake news on social media. By looking at data from Tweets during the 2016 US Presidential elections, where we saw significant Russian intervention, I identify the key defining characteristics of accounts that regularly post fake news. These characteristics can be used to identify these accounts, and subsequent action can be taken to ensure that are no longer able to spread misinformation. It is my belief that stopping the

spread of fake news at its source – the accounts actually spreading it – will greatly limit its effectiveness, and help people focus on real information, actual news.