Cornell University Project X Research Proposal

Alexander Wang
Cornell University
Ithaca, NY 14853
aw576@cornell.edu

Jerry Sun
Cornell University
Ithaca, NY 14853
js2746@cornell.edu

Kevin Zhou
Cornell University
Ithaca, NY 14853
klz23@cornell.edu

Kaitlyn Chen
Cornell University
Ithaca, NY 14853
kjc256@cornell.edu

Chenxin Fang
Cornell University
Ithaca, NY 14853
cf348@cornell.edu

Edward Gu
Cornell University
Ithaca, NY 14853
elg227@cornell.edu

1 Introduction

Even after half a year from when the vaccine was first released, not only are a little under half the US population fully vaccinated but Covid-19 is still raging on just as strong as before. A few reasons can explain why such infection numbers prevail such as emerging strains like the Delta variant. However, another reason that makes constant headlines every week is the spread of fake news and misinformation about the vaccine and also alternative medicine treatments that supposedly cure Covid-19. The narrative shifts on misinformation have changed from Covid-19 being a hoax to misinformation about the vaccine and marketing other, often harmful, treatments that are not approved to treat Covid-19. One thing that hasn't changed is the prevalence of such news circulating. The spread of Covid-19 through misinformation, thus pertaining to epidemiology, is our topic of interest.

2 Research Proposal

The primary question we have is can we determine the impact of fake news on the spread of Covid-19, identify the sources, and possibly flag it with reasonably high accuracy? Our proposal is to execute this in two main parts: one is to identify the impact and the spread of such misinformation and how much damage it's actually causing and the other is to identify such sources and flag them as such. We will be focusing on the ability to flag videos of misinformation rather than text analysis.

3 Background/Literature Review

There has been significant work in the area of fake news detection,[1] and recently with the pandemic, there has been some dedicated research into this area as well[2]. These papers show the expansion of fake news detectors in alleviating these concerns. Various extensions on this broad

general scope have been made as well such as detecting user query keywords that might trigger fake news results to show up[3]. Another area of exploration was specifically using this on different platforms[4] and in this case specifically Twitter. As a relatively new topic area, there are still advancements and areas to apply this research to.

4 Methodology

The methodology behind this research will be to scrape data collected from https://github.com/echen102/COVID-19-TweetIDs where post IDs are formatted and to find short video clips to classify and train on. There has been past research on the use of CNNs to classify videos[5] but none have tackled this in the context of Covid-19 and fake news. Videos specifically are a very popular format for information to spread as it is much more fruitful to persuade people through a video with demonstrations and visual tactics along with audio rather than just text which is also more susceptible to being flagged down.

5 References

[1] Zhang, Jiawei, et al. "FAKEDETECTOR: Effective Fake News Detection with Deep Diffusive Neural Network." ArXiv.org, 10 Aug. 2019, arxiv.org/abs/1805.08751.
[2] Mazzeo, Valeria, et al. "Detection of Fake News on COVID-19 on Web Search Engines." Frontiers, Frontiers, 1 Jan. 1AD, www.frontiersin.org/articles/10.3389/fphy.2021.685730/full.
[3] Shams, Abdullah Bin, et al. "Web Search Engine Misinformation Notifier Extension (SEMiNExt): A Machine Learning Based Approach during COVID-19 Pandemic." Healthcare (Basel, Switzerland), MDPI, 3 Feb. 2021, www.ncbi.nlm.nih.gov/pmc/articles/PMC7913172/.
[4] Glazkova, Anna, et al. "g2tmn At Constraint@AAAI2021: Exploiting CT-BERT and Ensembling Learning for COVID-19 Fake News Detection." ArXiv.org, 13 Jan. 2021, arxiv.org/abs/2012.11967.
[5] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar and L. Fei-Fei, "Large-Scale Video Classification with Convolutional Neural Networks," 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1725-1732, doi: 10.1109/CVPR.2014.223.

6 Backup Proposal

In our discussion of how misinformation—a purely social phenomenon—affects the physical health of entire populations, we similarly began comparing the policies of various countries in combating Covid-19—whether it be lockdowns, mask mandates, vaccination mandates, or social distancing policies. This led to the question: how does Covid-19 infection rate correlate with the public health policies implemented at a national level? Consequently, we have formulated this as our secondary project proposal with the end goal of creating a predictive model that could estimate a region's infection rate based on the public health policies currently in place.

References for Backup Proposal:
[6] Nussbaumer-Streit B, Mayr V, Dobrescu AI, Chapman A, Persad E, Klerings I, Wagner G, Siebert U, Ledinger D, Zachariah C, Gartlehner G. Quarantine alone or in combination with other public health measures to

control COVID‑19: a rapid review. Cochrane Database of Systematic Reviews 2020, Issue 9. Art. No.: CD013574. DOI: 10.1002/14651858.CD013574.pub2. Accessed 17 September 2021.

[7] Dandekar, Raj, and George Barbastathis. "Quantifying the Effect of Quarantine Control IN Covid-19 Infectious Spread Using Machine Learning." 2020, doi:10.1101/2020.04.03.20052084.