A.I. TO CATER RELIABLE NEWS SOURCES

(Technical Paper)

MISINFORMATION IN GOVERNMENT AND SOCIETY

(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

Samer Nahed Kadih

Fall 2021

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

Briana Morrison, Department of Computer Science

Catherine D. Baritaud, Department of Engineering and Society

Disinformation is false information deliberately meant to mislead. The subject of "fake news," and social media's role in exacerbating its effects, has been center-stage in political discussion since the 2016 U.S. presidential election (Tam, 2017). More recently, according to the Congressional Research Service, social media plays an active role in hampering efforts to mitigate the spread of the COVID-19 virus (Gallo & Cho, 2021, pp. 14-18). In order to keep up with the volume of data disseminating over the internet, machine learning models for natural language processing are typically employed for classifying the credibility of information (Asr & Taboada, 2019, p. 5). A variety of reasons may contribute to the general public's unwillingness to screen the news presented to them for bogus statements, such as not making time in their busy lives or not feeling confident about their technical literacy (Horrigan). The technical project will focus on building and training a machine learning model with the purpose of dynamically catering reliable news to users on social media. The STS research project will employ actor-network theory to better understand the motivations and processes involved in peddling misinformation in the national political network. These tightly coupled projects will work in tandem to provide a framework for social media engineers to responsibly deal with misinformation on their respective platforms.

Research Schedule

The research will be conducted over the course of next semester. The first two weeks will involve working on gathering data for the technical portion of the research project, as well as case studies for the STS portion–preferably related to the COVID-19 pandemic. The next six weeks will be technical in nature for both portions of the research, wherein models and networks will be designed based on present research in natural language processing and misinformation peddling. The models will be tuned appropriately to the specifics of the problem. The remainder

of the semester will involve building an interface to interact with the model in a practical sense, such as a single-page application or application programming interface (API), as well as enumerating how it may complement the STS portion in combatting fake news in standard social media pipelines.

NATURAL LANGUAGE PROCESSING TO CATER RELIABLE NEWS

Fake news is not a novel concept; even in 18th-century London, academics noted that "stories or gossip made it into newspapers which had just began to circulate among a broad public" (Asr & Taboada, 2019, p. 2). Today, the development of the internet amplifies the effects of this malpractice to a global scale. The human tendency to surround oneself with the familiar is the impetus for echo chambers: a phenomenon driven by human bias to consume content relevant to their platform, and avoiding any ideas that challenge their pre-existing beliefs (pp. 3-4). Corrupt actors in social networks, typically motivated by economic or political means, exploit this phenomenon by planting sensational posts on social media and having them organically spread throughout the platforms (p. 4). The speed at which a non-sensical idea can reach the masses is frightening, but even more frightening are the lasting consequences: "people tend to remember facts and events that have been repeatedly mentioned, even when the repeated mention is in the context of a retraction or myth debunking" (p. 4). Detecting misinformation in news sources is an integral part of helping regular consumers of social media responsibly navigate the political environment. Because of this, social media engineers have a responsibility to programmatically detect fake news before it is able to spread throughout.

There is a plethora of well-documented, rudimentary models for detecting fake news (Asr & Taboada, 2019, p. 5). A modern method is adaptive graph convolutional networks, in which various types of information, such as visual and text, have connections whose properties can be

leveraged to better classify that information (Qian, Hu, Fang, & Xu, 2021, pp. 7-11). A sophisticated model proposed by Ruchansky et al. splits the learning into different modules, as depicted by Figure 1. The "capture" module uses a recurrent neural network (RNN) to characterize the temporal patterns present in the distribution of an article. The "score" module attempts to characterize user behavior when engaged with an article. Finally, the "integrate" module attempts to construct a relationship between the "capture" and "score" modules to make a conclusion on the veracity of that article. There is also research being done on maximizing influence in social media channels by distributing information in a decidedly fairer way, so that demographic biases are minimized in the spread of that information in a social network (Stoica & Chaintreau, 2019, p. 1).

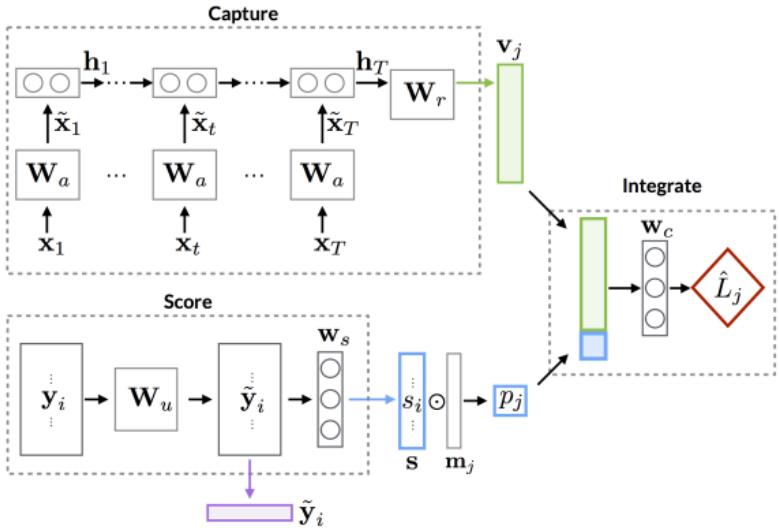

Figure 1: CSI model specification. This figure visualizes the different aspects of the Capture Score Integrate model. (Adapted by Ruchansky, N., Seo, S., Liu, Y. (2017). ACM).

Although these algorithms work well to detect misinformation or classify network efficiency, they consequently take on a passive role in combatting misinformation. In combining these technologies, the technical project seeks to *actively* combat the spread of fake news by building a model to capture the present state of the political environment and deliberately *cater*

reliable news sources to the user base accordingly. Over the course of a semester capstone class, the model, trained on public datasets tabulating news and their classified credibility, will be built to take as input the network attached to a user, and jointly considering the present political environment, output a list of credible news sources accordingly.

Findings will be laid out in a scholarly article, where a critical element to the success for this project is in gathering a large pool of quality data regarding misinformation in social media, so that the machine learning models are trained with a non-trivial set of data for practical applications. Then, a plethora of different flavors of models, such as predicative modeling or multi-modal representations, will be tested and tuned against readily available benchmarks. Finally, it is pivotal that the interface for this project can access a pool of news sources which can be catered to the end-user, such as by a third-party news API, after the news that ought to be catered to the user is properly identified. A diverse palette of news will be likely be suggested to minimize bias. A proper model will classify user-typed or viewed social media posts as either true or false. If news is classified as the latter, then, either by the same or a different model, derive which articles from verified news sources in recent history are related to the context of the accused news. News from verified sources is considered better suited for conveying information to unsuspecting consumers of social media with regards to the present political landscape.

IMPACT OF MISINFORMATION ON GOVERNMENT AND SOCIETY

The STS project has to do with the propagation of news in social media settings. Specifically, analyzing different models for *how* and *why* misinformation propagates over media channels from a technological perspective and its impact on government and society. One of the U.S. government's chief subjects of interest involve mitigating media channels' role in exacerbating misinformation (Gallo & Cho, 2021, pp. 14-18). Particularly, with respect to the

COVID-19 virus, whose subject is most relevant to the present political environment (Gradoń et al., 2021, pp. 1-4). How this information propagates, resulting in polarizing ideas in social media, can be modeled by immediate constituent analysis: a profound method for analyzing complex networks (Prasetya & Murata, 2020, p. 7). To anchor this model to a problem that is still actively being a detriment to society, the Biden administration's efforts to increase vaccination rates was met with hesitation by the public, and a consequent surge in hospitalized patients (Ivory et al., 2021). This is likely a cause of targeted peddling by corrupt actors in the network. If a suitable model can be determined for how the dynamics of our present political network behave, then a method for minimizing the effects of fake news propagation can be employed.

This project aims to invoke the use of actor-network theory (ANT) to the particular case of misinformation in the political network. ANT is especially useful as a method for modelling the dissemination of fake news, since it seeks to highlight the motives of adversarial actors. Different actors involved in the network have different influences in how misinformation is spread, each with their own psychological trends and oddities that are outlined in present-day research. The general trends of a network can be described by the probabilistic laws governing the network. Although ANT cannot feasibly predict the actions of any one actor, over a large sample these populations can be understood to act in a certain way with some level of confidence, and thus can be modeled by stochastic processes (Venezuela et al., 2019, pp. 808-810). Figure 2 demonstrates how Bayesian statistics can be used to fit parameters to how a network operates. When compared to other models proposed by Venezuela et al., it appears that politically actives citizens in a network tend to distribute misinformation more often than the less engaged. By jointly using ANT to study the motivations of different actors, as well as breaking

down the actors into measurable stochastic processes, there are then two different avenues of approaching the problem of minimizing the dissemination of fake news. It can either be done between the relationships of different actors, or on a more granular level.

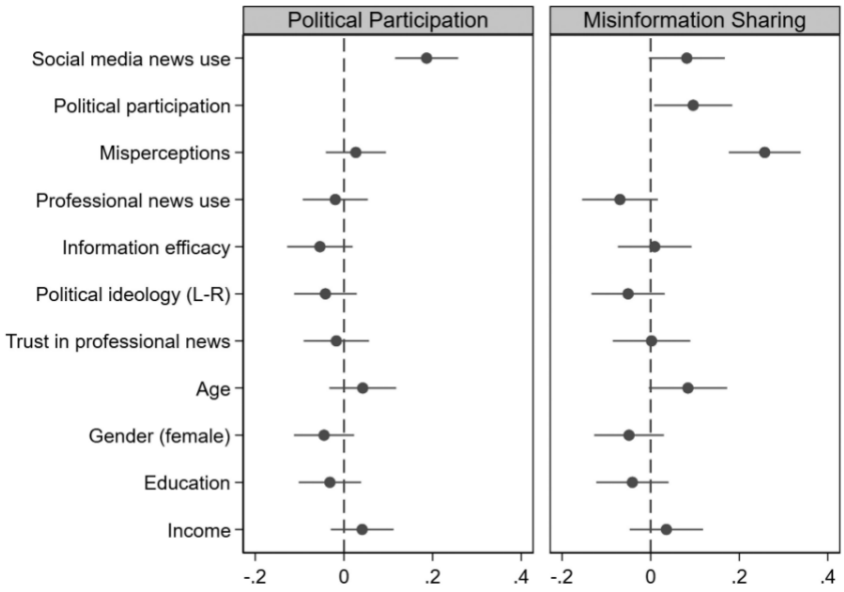

Figure 2: Standardized coefficients. This figure visualizes the estimated magnitude and spread of network parameters between political participation and misinformation sharing. (Adapted by Valenzuela, S., Halpern, D., Katz, J. E., & Miranda, J. P. (2019). Digital Journalism 2019).

Findings will be laid out in a scholarly article, where a different way of framing the national political network will be suggested so that misinformation dissemination is minimized. A model will have to be constructed for how the major actors in the network behave, including relevant statistical relationships between them. For example: given that an actor disseminates fake news, how does its retention rate or spread compare to that of verifiable news? Figure 3 highlights a suggested model for information dissemination. This simple model makes a few assumptions, or hypotheses, about our political environment to explain its dynamics accordingly (Valenzuela et al., 2019, pp. 806-807). Hypothesis 1 (H1) is that social media news use is positively correlated with political participation. Hypothesis 2 (H2) is that political participation

is positively correlated with the spread of misinformation. Hypothesis 3 (H3) is that misperceptions about the news cycle will moderate that relationship between political participation and the sharing of misinformation in a network so that the relationship is stronger for misinformed users and weaker for informed users.

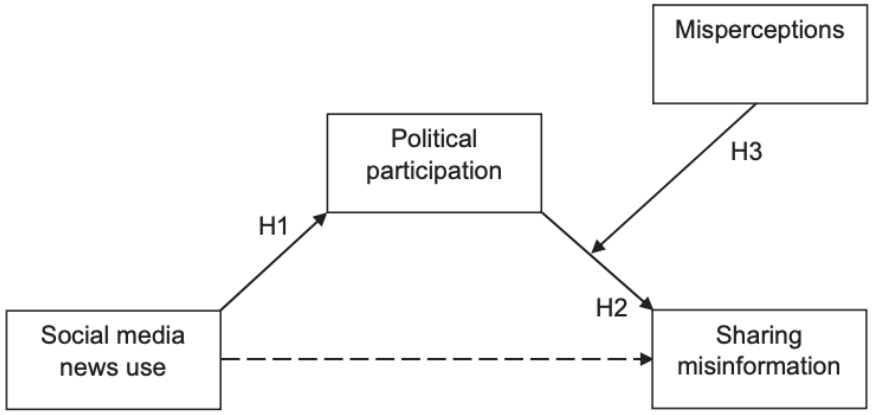

Figure 3: Model of information dissemination. This figure visualizes the interactions between social media and political participants in propagating fake news. (Adapted by Valenzuela, S., Halpern, D., Katz, J. E., & Miranda, J. P. (2019). Digital Journalism 2019).

ANALYSIS AND REMEDY OF MISINFORMATION NETWORK

The rise of easy-access news on social media makes it harder for the average American to gauge its credibility, since people do not typically dedicate the amount of time necessary to successfully classify news for themselves (Horrigan). While the STS topic seeks to answer *how* misinformation spreads, the technical topic deals with the process for *how to mitigate* the spread of misinformation. Thus, the technical project will focus on building a model to automatically and intelligently cater news related to a network's interests, while the STS project will focus on understanding the psychology behind the spread of fake news and its effects on public policy, so that in tandem these findings can be used to determine a good framework for disseminating news on social media platforms so that the effects of misinformation on the state of the political network is minimized.

References

Asr, F. T., & Taboada, M. (2019). Big data and quality data for fake news and misinformation detection. *Big Data & Society.* https://doi.org/10.1177/2053951719843310

Chan, M. S., Jones, C. R., Jamieson, K. H., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science, 28(11), 1531-1546.* https://doi.org/10.1177/0956797617714579

Gallo, J. A., & Cho, C. Y. (2021). Social media: Misinformation and content moderation issues for Congress. *Congressional Research Service*. https://crsreports.congress.gov/product/pdf/R/R46662

Gradoń, K. T., Hołyst, J. A., Moy, W. R., Sienkiewicz, J., & Suchecki, K. (2021). Countering misinformation: A multidisciplinary approach. *Big Data & Society, 8(1).* https://doi.org/10.1177/20539517211013848

Horrigan, J. B. (2017). How People Approach Facts and Information. *Pew Research Center*. https://www.pewresearch.org/internet/2017/09/11/how-people-approach-facts-and-information/

Ivory, D., Leatherby, L., Gebeloff, R. (2021). Daily report: Least vaccinated U.S. counties have something in common: Trump voters. *The New York Times*. https://www.nytimes.com/interactive/2021/04/17/us/vaccine-hesitancy-politics.html

Prasetya, H. A., & Murata, T. (2020). A model of opinion and propagation structure polarization in social media. *Computational Social Networks, 7(1).* https://doi.org/10.1186/s40649-019-0076-z

Qian, S., Hu, J., Fang, Q., & Xu, C. (2021). Knowledge-aware multi-modal adaptive graph convolutional networks for fake news detection. *ACM Transactions on Multimedia Computing, Communications, and Applications, 17(3), 1–23.* https://doi.org/10.1145/3451215

Ruchansky, N., Seo, S., Liu, Y. (2017). CSI: A hybrid deep model for fake news detection. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 797-806.* https://doi.org/10.1145/3132847.3132877

Stoica, A.-A., & Chaintreau, A. (2019, May 13). Fairness in social influence maximization. companion proceedings of the 2019 World Wide Web Conference. *WWW '19: The Web Conference.* https://doi.org/10.1145/3308560.3317588

Tam, P. (2017). Daily report: The continued creation and dissemination of fake news. *The New York Times*. https://www.nytimes.com/2017/01/19/technology/daily-report-the-continued-creation-and-dissemination-of-fake-news.html

Valenzuela, S., Halpern, D., Katz, J. E., & Miranda, J. P. (2019). The paradox of participation versus misinformation: Social media, political engagement, and the spread of misinformation. *Digital Journalism, 7(6), 802-823.* https://doi.org/10.1080/21670811.2019.1623701