Evaluation of Solutions for the Problem of False Information Prevention and Future Directions

A Research Paper 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

Eric Stoloff
Spring 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

Signature________Eric Stoloff________. Date___4/1/2021___.
Author Name
Approved__________________________. Date______________.
Richard Jacques, Executive Lecturer, Department of Science, Technology, and Society

Introduction

The spread of false information has become a major problem in our society today. Although mainstream news networks are not without incident, social media is the most important medium through which damaging false information propagates. Social media is defined as "any online or digital medium provided and/or collected through a channel that enables the two-way sharing of information, involving multiple parties." (DHS, 2018). In 2016, social media was the most important information source for 51% of people (Newman et al., 2016), and its popularity and use continue to grow every year (Chaffey, 2020). This growing user base presents a growing opportunity for the spread of false information because social media allows people to not just receive information from news and other networks, but the ability to create their own media and the means to potentially spread it worldwide (Dafonte-Gomez, 2018). Although social media and the internet are relatively young technologies, there are already a multitude of examples that illustrate the consequences of the proliferation of false information on social media, including the 2016 American Presidential Election (Allcott & Glentzkow, 2017). In the three months before the election, there were over 156 news stories that circulated on social media that were determined to be false, the vast majority of which favored Donald Trump (Allcott & Glentzkow, 2017). Not only were there more instances of false information biased towards Trump, but these stories were shared over four times more than false information supporting Hillary Clinton. Although it is possible for an individual or group to create false information that serves a purpose, the case of the 2016 election can be explained with the advertisement revenue that the creators of many of the false stories received (Allcott & Glentzkow, 2017). After initial circulation, these stories were correctly identified as fake and flagged, but the major problem in preventing the spread of false information is the scope and scale of it. There are now 3.96 billion social media users and 4.57 billion total users of the internet worldwide (Chaffey, 2020). The issue with finding cascades of false information is that all the content generated from these users is an impossibly large pool of data that would take incredible amounts of computing power to parse through. In addition, about 90% of information available is credible, so it is easy for false information to naturally hide (Kumar & Shah, 2018). Solutions that will expose false information will need to be effective while also being practical enough to be implemented within real computer networks. The final factor in determining how favorable a possible solution will be is the ethical implications surrounding what defines false information and the level of intrusion into personal information that may occur in the process of finding false information. Given that people have personal liberties like Freedom of Speech in the U.S., there is some question as to whether the Government or social media networks have the right to prevent people from sharing through social media (Maitland & Lynch, 2020). In this paper, I evaluate potential solutions to hinder the spread of false information with these three main factors as the criteria to judge them. I will explore the behavior of users and their interaction with information on social media in order to gain insight into how potential solutions address this and then evaluate a variety of these solutions. False information, especially that which is sneakily presented as fact, is too damaging to go unchecked and I identify how it can be addressed effectively and ethically.

Part 1: Users' Interaction with Information and Its Effects

The expansion of false information through social media is a major problem in modern times with social media continuing to expand annually. As users' capacity to participate in the dissemination of information has increased due to social media, they have stepped up to fill it (Dafonte-Gomez, 2018; Westerman & Spence, 2014). In doing so, users may take part in the spread of false information. In this context, false information is defined as information or a story propagating through social media which is ultimately verified as being false or inaccurate (Guo et. al, 2020). It is an umbrella term which includes the many types listed in Figure 1 below.

Term	**Definition**
Rumor	"An item of circulating information whose veracity status is yet to be verified at the time of posting"
Fake News	"A news article that is intentionally and verifiably false"
Hoax	"A deliberately fabricated falsehood made to masquerade as truth"
Click-bait	"A piece of low-quality journalism which is intended to attract traffic and monetize via advertising revenue"
Disinformation	"Fake or inaccurate information which is intentionally false and deliberately spread"
Misinformation	"Fake or inaccurate information which is unintentionally spread"

Figure 1: Definitions of various types of false information (Guo et al., 2020)

Users share content on social media for the simple fact that humans love sharing information. We would not be where we are as a species today without our incessant urge to communicate with each other and socialize. Regardless of the content, the human brain produces reward stimuli for the act of sharing information (Dafonte-Gomez, 2018). This is critical to the spread of false information because a user unintentionally sharing it will still allow them to experience a feeling of satisfaction, however minute. The motivations for users to share on social media most commonly include entertaining an audience, trying to change the opinions of others, informing people about something, and creating shared experiences and a feeling of belonging (Ardevol-Abreu et al., 2020). Save entertainment, these rationales all come with the expectation that the sender is giving the recipient valuable information, which also comes with the expectation of truth. It is when the sender has misinformed the recipient that the problem occurs, because if the recipient had the expectation of truth then they are more likely to consider the information being presented as truth. Most people assume that everyone else has generally good intentions and would not try to dupe them, so they tend to believe what people say. Research has even shown that in order for someone to comprehend a piece of information, they need to believe it as true, even if only temporarily (Lewandowski et al., 2012). This problem is one of the root causes of the spread of misinformation. Instead of being vigilant fact-checkers, social media users lazily skim through information and interact with social media by "snacking." This is "a practice that is not about pursuing in-depth knowledge or developed opinions, but about diversion: users consume bits and pieces of information in a relaxed, easy-going fashion to gain a sense of what is going on.” (Dafonte-Gomez, 2018). The motivations of users to share information on social media combined with the complacency of their recipients is one of the mechanisms that false information spreads, but it is far from the only one. The second issue relating to user behavior has to do with users' previous beliefs, values, and perceptions of the world around them when they encounter a piece of false information. People like to avoid conflict, and they enjoy interacting with media that fits their current worldview. Research suggests that information has a higher likelihood of being accepted as truth when it fits in with other information users believe, and information that doesn't fit in results in negative emotions and a decreased ability to process the information (Lewandowsky et al., 2012). This culminates in a higher cost on the user to accept information that challenges their current views. Consequently, users like to enjoy content that does not require this extra effort and goes along with their beliefs and viewpoints. This phenomenon is known as selective exposure and there is significant research suggesting that it is a dominant force in social media users' consumption of media (Cinelli et al., 2020). Selective exposure allows users to never have to confront truth, and this problem is multiplied by their interactions with other users. Multiple users with like-minded beliefs can further be shielded from inflammatory content because they can reinforce each other's beliefs. This leads to the creation of echo chambers, where the same set of ideas and values are supported by everyone, leading them to see an altered version of the truth (Dafonte-Gomez, 2018; Kumar & Shah, 2018). The divisions in social media users on politics best illustrate this, where the user interactions within political groups dwarfs user interactions between groups.

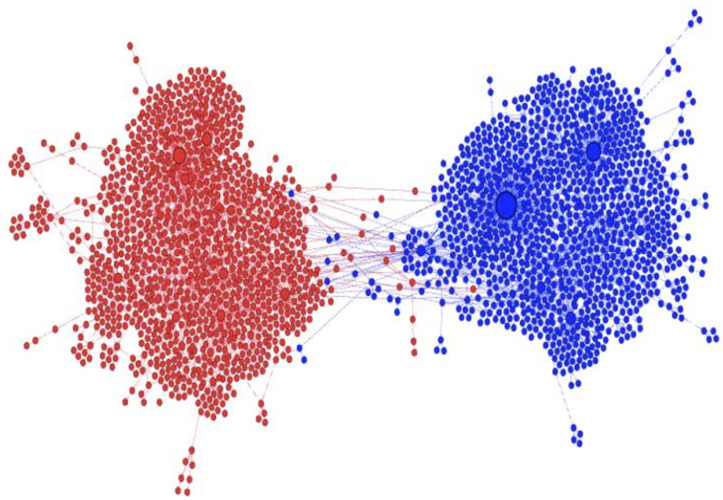


Figure 2: Graph of retweets on a controversial #beefban topic shows the formation of echo chambers, from Kumar & Shah, 2018. Red dots are used to represent conservative users while blue dots are used to represent liberal users.

The formation of echo chambers shows that important to users' beliefs and views are users' perceptions of other users and their beliefs. People tend to trust in the social consensus and although humans have come a long way, we are still subject to the primitive rule of safety in numbers. In an echo chamber, the repetition of false information can actually create a divide between what the social consensus is in reality and what the members of the echo chamber think the social consensus is. This can be demonstrated by a study of Australians' attitudes towards asylum seekers, where just 1.8% of people surveyed had strong negative attitudes towards asylum seekers, but yet that small group thought that 69% of all Australians and 79% of their friends held the same beliefs (Pedersen, Griffiths, & Watt, 2008). In this case, the minority proved that they believe in an actual falsity and this presents the opportunity down the road to make a bad decision based on that false belief. In this case, it could be attempting an act of violence against asylum seekers believing that they have the support of most of Australia. If the formation of echo chambers on social media was not bad enough due to other factors already, the algorithms that sites like Facebook and Twitter use to recommend media to users actively contribute to their creation (Ardevol-Abreu et al., 2020; Dafonte-Gomez, 2018). These algorithms recommend media that they think the user will like based on their previous likes and dislikes in order to retain the user's attention for the longest possible time and keep them engaged so that the companies can maximize their advertising revenue. This furthers the level of selective exposure which ultimately grows echo chambers and enables the spread of more false information within them. The behavior of users and the environment they inhabit on social media is fraught with problems that allow false information to propagate. When people believe something erroneous, they put themselves in a future position to make a brash decision based on that belief, and people are put in this position too often in society today. In addition, the initial evidence suggesting that fake news can successfully modify people's unconscious behavior is worrying (Bastick, 2021). It is impossible to change the behavior of all users directly, so social media companies must find a way to force adoption of new behaviors that prevent false information from taking root in the user base. The next section delves into potential solutions to the problem and analyzes them.

Part 2: Evaluation of Reactive Approaches to False Information Remission

As previously mentioned, there are many types of false information. This paper covers methods to finding them in a broad sense, especially because there is significant overlap between the different types. However, rumors and click-bait are not the focus of this paper because they are not necessarily malicious and are loosely defined. Almost anything said on social media could be classified as a rumor, or piece of unverified information, including sarcasm. In addition, click-bait in and of itself is not dangerous, but it becomes dangerous when the content of the click-bait falls into one of the other types of false information. There are also many proposed solutions to preventing the spread of false information on social media. These can be divided into two groups: active and reactive. Active approaches work by altering the way in which users interact with information and each other on social media and deliberately change the way that information is spread in the network. Reactive methods of prevention work by responding to the flow and content of information on the network and correcting the flawed information itself. This section focuses on reactive methods of prevention and evaluates them in terms of their effectiveness in identifying and neutralizing false information, practicality and scalability, how intrusive into user privacy the method is, and the time it takes for the solution to work. The first of these is machine and deep learning-based approaches that make use of content, metadata, and user-interaction data in order to find false information in the social media sphere (de Beer & Matthee, 2020). For this category of solution, I examine the machine learning method proposed and tested by Atoderesei, Tanaselea, and Iftene in their 2018 paper because of its success and its use of not only content, but metadata and user data (Atoderesei, Tanaselea, & Iftene, 2018). Additionally, Facebook currently uses an approach to detecting false information which, like the approach being examined, places a high value on the popularity of a post, making this an even more important algorithm to consider (Atoderesei, Tanaselea, & Iftene, 2018). The algorithm works by using a Twitter crawler to add Tweets to its database. Then, when the application is used to check the validity of a tweet, it uses Named Entity Recognition to separate the Tweet into its entities and sentiments which get compared with other tweets in the database. The degree of similarity between the tweet in question and the other tweets and the number of other tweets similar to it are two of the most important factors for this algorithm. The final one is the user score that the algorithm develops for each user. User scores start at 0 and increase for the amount and validity of the user's tweets (Atoderesei, Tanaselea, & Iftene, 2018). Collectively, these factors place a high value on the popularity of a tweet. The algorithm does not classify tweets binarily, but assigns them a tweet score that is a function of the three previously mentioned factors, with firm classifications at both ends of the spectrum and the lack of a result in the middle (Atoderesei, Tanaselea, & Iftene, 2018). Building a large database of tweets to compare with, the professors who carried out this project were able to demonstrate that their application was able to identify false information accurately and establish a confidence measure of how likely the information was to be false. In their paper, they used various use cases scattered on the spectrum of truth, along with multiple users, to show that the intended result was achieved (Atoderesei, Tanaselea, & Iftene, 2018). Although the authors admit that using the popularity of a tweet may not be the best way to ascertain a tweet's truthfulness, they still use the results of their algorithm as the sole metric with which to judge their method by. This is a glaring error in terms of ethics towards the users of social media networks, as well as a miscalculation towards the nature of how false information spreads. The idea of each user having a user score raises some concern that users could be treated unfairly based on their user score (Maitland & Lynch, 2020). Because user score is factored into the calculation of the veracity of a tweet, users with a low score will consistently have their tweets return a lower confidence of being true compared to users with a higher score. Although the professors describe the use case of the same tweet for multiple users returning the same confidence score, they fail to mention the user scores that were used, so it is unclear how a large gap in user score affects the results (Atoderesei, Tanaselea, & Iftene, 2018). They also used a tweet about the Super Bowl for this case that most people could quickly agree is true. The other element that this solution overlooks is that false information spreads faster and to more users than true information (Vosoghi, Roy, & Aral, 2018). By using popularity as the main driver behind the algorithm's classification of a tweet's authenticity, it opens the door for the small percentage of false information that does go viral to evade detection (Kumar & Shah, 2018). These inciteful posts spread quickly to thousands of users and it is no wonder that, even with numerous attempted algorithm improvements by Facebook, at least six of the ten top daily news stories on that platform with the most user interaction come from right-wing opinion sites and influencers (Bauerlein & Jeffery, 2021). Another reactive method of countering the spread of false information is fact checking. Information can be manually checked by experts or by individuals in the privacy of their own home, but many social media sites have recently launched fact-checking initiatives including one that Facebook launched in the wake of the 2016 presidential election with a variety of fact-checking sites like snopes.com (Ardevol-Abreu et al., 2020; Dafonte-Gomez, 2018). This method is a step in the right direction; however, it still has its downsides. The first problem that arises with fact-checking is trust. The below graph illustrates the slow decline in media trust that has occurred, with the change being more extreme for republicans.

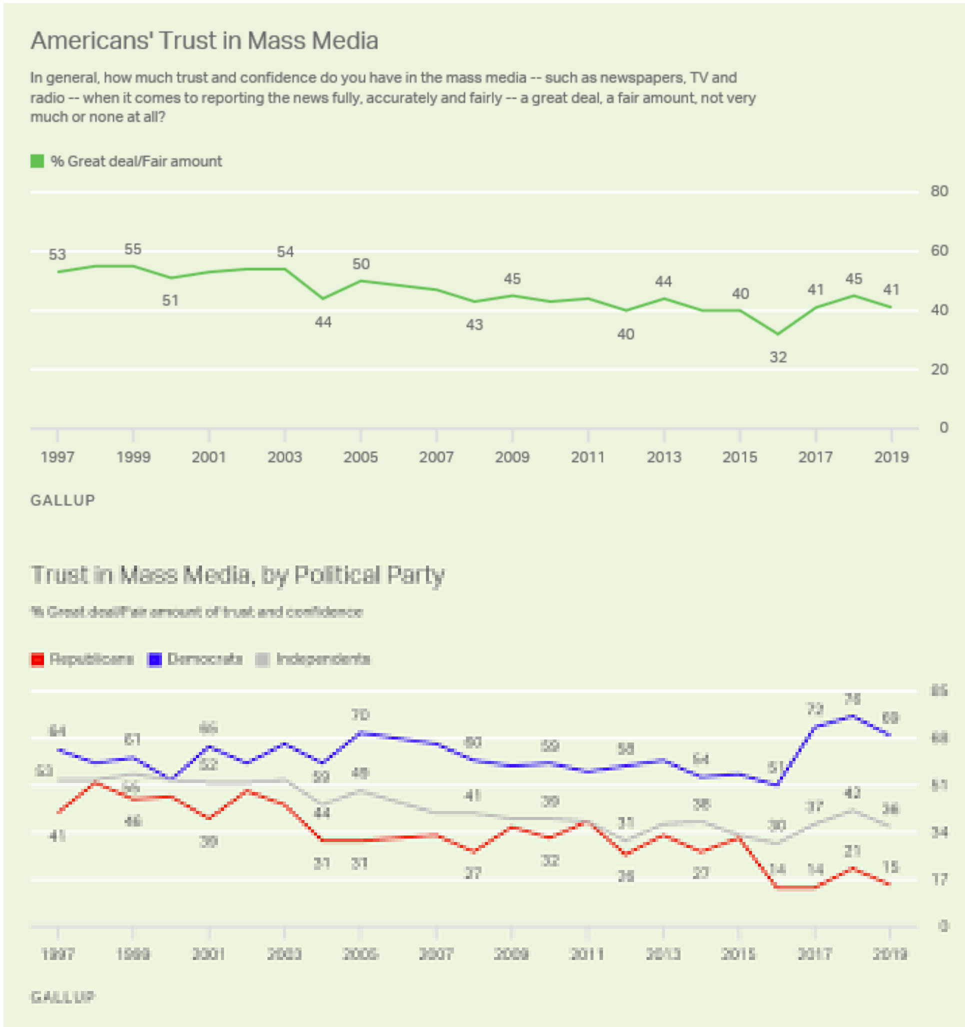


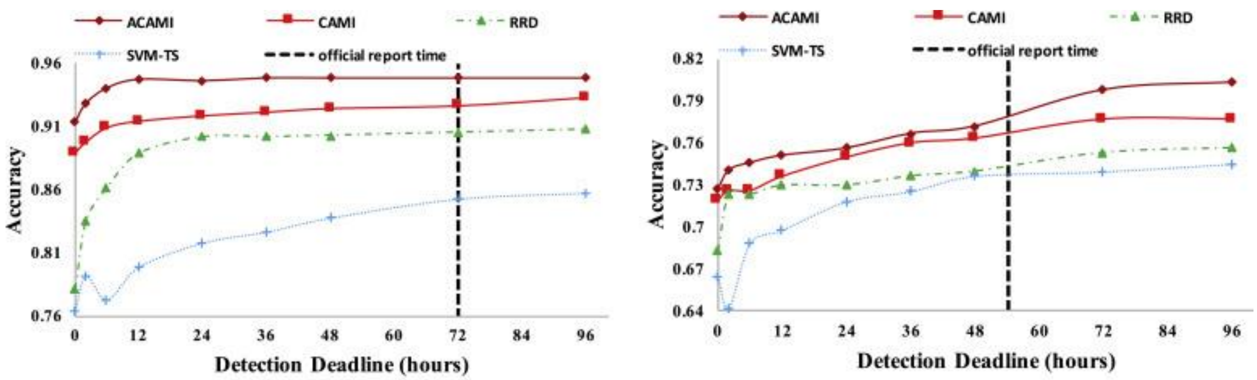
Figure 3: Graphs of media trust in general and by political party (Brenan, 2019)

With users' trust in social media falling, their trust in social media fact checkers is falling as well (Ardevol-Abreu et al., 2020). This means that users will completely disregard fact-checking measures if they don't believe they can be trusted. An additional problem with fact checkers occurs because of the cognitive biases of confirmation bias and disconfirmation bias. Essentially, users tend to favor arguments that fall into their narrative while ignoring contradictory information. (Ardevol-Abreu et al., 2020; Dafonte-Gomez, 2018). These lead users to readily accept information they agree and feel comfortable with, while rejecting information that does not, regardless of fact-checking and corrective measures (Ardevol-Abreu et al., 2020; Chou, Gaysynsky, & Vanderpool, 2021). Yet another concern for fact checking is that it does not consider the emotional response that the information elicits in the user. It has been shown that arousal can affect the way that users process misinformation as well as the likelihood that users will share information (Chou, Gaysynsky, & Vanderpool, 2021; Lewandowsky et al., 2012). A study by professors at the University of La Laguna in Tenerife concluded that people do not perceive the results of Facebook fact checkers as an important factor in deciding the falsity of information. The study was conducted across a politically and demographically diverse group and yet it found that the fact-checkers and associated warning messages had a minimal influence on the participants' decisions (Ardevol-Abreu et al., 2020). Not only does this study illustrate the pitfalls of fact-checking, but it is a microcosm of the pitfalls of reactive methods of false information prevention in general. This is because of the ineffectiveness of correcting false information after users engage with it. In fact, the only proven method to improve the effect retractions of misinformation have on the user is when the retraction fills the coherence gap in the users' understanding of information (Lewandowsky et al., 2012). This ability is not present in the machine learning solution or fact checking, which fail to provide a unique response to every piece of false information simply due to the scale of social media. Both of these methods also take time to work, with a manual component being required from users in both cases, and this is simply not fast enough for social media. False information takes an average of 12 hours to be debunked and in that time, it has the potential to go viral (Kumar & Shah, 2018). From the criteria used, it is increasingly clear that reactive approaches to mitigating the spread of false information still have a lot of issues. In the next section, I shift toward a discussion of active approaches and how they fit the evaluation criteria.

Part 3: Evaluation of Active Approaches to False Information Remission

In this section I evaluate active approaches to preventing the proliferation of false information on social media. Active approaches use social engineering to change how users interact with information on a behavioral and mental level and they work at the same time-scale of users sharing information. The first method that I examine is using preemptive inoculation in order to weaken a user's susceptibility to false information. This works much like putting a weakened copy of a virus inside the human body, by presenting a weak version of a challenge in order to spark critical thinking (van der Linden, Roozenbeek, & Compton, 2020). This has been proven effective in mitigating the effects of exposure to false information, as a study using false information about climate change showed (van der Linden et al., 2017). The solution examined for this case is the paper by Cambridge professors Sander van der Linden and Jon Roozenbeek and Dartmouth professor Josh Compton on active inoculation. Active inoculation works by warning and making users aware of the manipulation techniques that false information employs instead of inoculating about individual topics like vaccination or QAnon (van der Linden, Roozenbeek, & Compton, 2020). In their paper, the authors endorse implementations of active inoculation like the online game *Bad News*. In this game, players take on the role of a false information spreader and learn about manipulation techniques like fear mongering and fake experts, get warnings about misinformation, and see actual examples (van der Linden, Roozenbeek, & Compton, 2020). Studies involving *Bad News* demonstrate a dramatically increased ability of users to resist false information and an increased confidence in identifying it. The authors also advocate for another online game specifically related to false information about Covid-19 called *Go Viral!* which was developed by the U.K. and the WHO. This game focuses on teaching users how to resist manipulation techniques specifically related to Covid-19 (van der

Linden, Roozenbeek, & Compton, 2020). The effectiveness of this new game is still largely unproven, but it still uses active inoculation to address the manipulation techniques commonly found in false information. Although preemptive inoculation is promising in cutting down on the spread of false information because it can "immunize" users, the authors admit that their solution is not perfect and more research in this area is needed (van der Linden, Roozenbeek, & Compton, 2020). The first issue with this solution is the practicality of inoculating enough people to achieve online "herd immunity." With billions of social media users worldwide, it would be hard to practically force users to inoculate themselves by playing an online game (Chaffey, 2020). Inoculation would have to be widespread in all types of media in order to have an effect, and this problem is compounded by the "booster shots" required to maintain resistance to false information (van der Linden, Roozenbeek, & Compton, 2020). The online game suggested by the authors may not be practical, but active inoculation is definitely scalable into other forms of media. For example, TV commercials and streaming services both have good potential for inoculation to expand into, as the former was exhibited by the *Truth* campaign against the tobacco industry (Chou, Gaysynsky, & Vanderpool, 2021). This solution is also not a heavy burden with respect to time. Although playing an online game does require some time commitment, the effects of the inoculation last for months afterward (van der Linden, Roozenbeek, & Compton, 2020). The main upside is that this solution does not require a manual action by a user or fact checker in order to work. It improves the social media climate almost automatically, and future implementations of inoculation could be made to be even more seamlessly integrated into media to lessen the time commitment of users. A further benefit to an active inoculation solution is that it seems to present no real ethical concerns for users. There are virtually no downsides to playing an online game that educates about the techniques present in false information that pull users in. Plus, if inoculation efforts were more interspersed in many forms of media, the practice would then be more implicitly bound to users' everyday lives. This is due to the steady increase in the number of different media that people consume (Chaffey, 2020). The solution of active inoculation checks the majority of the criteria, and the only area in which it lacks is its effectiveness. Despite the improvements that it allows users to make in terms of their resistance to false information and manipulation techniques, it is certainly not perfect at catching every piece. The second active method of false information prevention I discuss is the early detection approach using attention-based convolutional neural networks (CNNs) proposed and tested by Feng Yu, Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan from the Center for Research on Intelligent Perception and Computing in the Chinese National Laboratory of Pattern Recognition. Although this method works using data collected from social media, it can still be considered an active approach due to its ability to work at the same time that information is spreading and prevent users from engaging with dangerous cascades of false information. This approach applies machine learning with CNNs actively, and uses them to analyze information cascades in real time in order to extract significant features and high-level interactions that unveil false information (Yu et al., 2019). Based on features extracted from the content and more importantly the way in which related information is shared in the network, the attention-based CNN can distinguish between which information cascades are harmless and which contain dangerous false information (Yu et al., 2019). This means that this method can learn how information is spreading in real-time and refine its judgements as it continues to spread. In order to test this method, they tested it on datasets from Twitter and Sina Weibo alongside other algorithms for identifying the spread of misinformation such as SVM-TS and RRD (Yu et al., 2019). They concluded that the attention-based convolutional approach to misinformation identification (ACAMI) achieves a higher degree of accuracy than the other algorithms and more importantly it converges on a high degree of accuracy quickly enough to be used for early detection (Yu et al., 2019).



(a) Weibo dataset (b) Twitter dataset

Figure 4: Comparison of ACAMI with other algorithms, plotting accuracy of identifying false information over the time of a particular information cascade or series of related posts (Yu et al., 2019)

Once again, because of the average 12-hour delay between false information's release and its debunking, the ACAMI method represents a significant improvement in the real-time detection of false information (Kumar & Shah, 2018; Yu et al., 2019). It seems to start converging on its maximum accuracy in about six hours for the Sina Weibo dataset and even less for the Twitter dataset, well within the 12 hours. The ACAMI method is also ethically favorable because of its ignorance of individual users. Although it uses users' posts and interactions on social media, the high-level conclusions it draws are not influenced by what users are involved in a particular information cascade, but rather how the users in the cascade interact (Yu et al., 2019). Therefore, it fails to retain any information which would result in the algorithm discriminating against certain users. Although the ACAMI solution has many upsides and shows promise, it does have drawbacks with its practicality and scalability. The datasets used in the testing are small compared with the breadth of social media, so it is unclear how a CNN can be applied to parse through this larger landscape (Chaffey, 2020; Yu et al., 2019). It is possible to use this tool more surgically, in order to monitor known echo chambers and hives for false information, but the amount of computing power it would take a CNN to oversee social media as a whole is simply too great.

Conclusion

The issue of false information propagating through social media cannot be understated for its past and potential negative effects on society. In this paper, I have evaluated four different solutions, both reactive and active, according to the criteria of how successful the solution is in preventing false information spread, how practical and scalable it is to implement for the whole of social media, the time it takes to work, and how ethically sound it is towards users. It has been shown with the deep-learning and fact-checking methods that although they are pretty successful, they are too slow to combat the fast spread of misinformation over social media. In addition, the deep-learning approach poses serious ethical risks with its use of a user score. More importantly, the main reason that these reactive methods are not the best choice for false information prevention is because of how ineffective corrections are for eliminating the effects of false information (Lewandowsky et al., 2012, Ardevol-Abreu et al., 2020). Coupled with the fact that active methods fulfilled more of the evaluation criteria, it is clear that they are the better route to go in the future. Of the two active methods previously discussed, active inoculation appears to be the better approach purely on the basis of practicality and scalability. It did have the potential issue of effectiveness as an online game but this can be improved with more widespread use and expansion into other media, which will increase repetition (Lewandowsky et al., 2012). If this criterion is more fully satisfied in the future, then with its scalability, lack of ethical concerns, and ability to work in real-time, active inoculation is the best method to use for false information prevention. One possible way this might be done in the future would be to amalgamate active inoculation with the user interface of social media applications.

Word Count: 4418

References

Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. Journal of Economic Perspectives, 31(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Ardèvol-Abreu, A., Delponti, P., Rodríguez-Wangüemert, C. (2020). "Intentional or inadvertent fake news sharing? Fact-checking warnings and users' interaction with social media content". Profesional de la información, v. 29, n. 5, e290507.

Atodiresei C-S, Tănăselea A, Iftene A. Identifying fake news and fake users on twitter. Procedia Comput. Sci. 2018;126:451–461. doi: 10.1016/j.procs.2018.07.279.

Bastick, Z. (2021). Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation. Computers in Human Behavior. 116. 106633. 10.1016/j.chb.2020.106633.

Bauerlein, M., & Jeffery, C. (2021). How Facebook Broke Us. Mother Jones, 46(2), 44–62.

Brenan, M. (2019). Americans' Trust in Mass Media Edges Down to 41%. Gallup Poll. https://news.gallup.com/poll/267047/americans-trust-mass-media-edges-down.aspx

Guo, B., Ding, Y., Yao, L., Liang, Y., & Yu, Z. (2020, August 1). The Future of False Information Detection on Social Media: New Perspectives and Trends. ACM Computing Surveys, 53(4), 68 - 103.

Chaffey, D. (2020, August). Global Social Media Research. Smart Insights. Retrieved from https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/

Chou, W.-Y. S., Gaysynsky, A., & Vanderpool, R. C. (2021). The COVID-19 Misinfodemic: Moving Beyond Fact-Checking. Health Education & Behavior, 48(1), 9–13. https://doi.org/10.1177/1090198120980675

Cinelli, M., Brugnoli, E., Schmidt, A. L., Zollo, F., Quattrociocchi, W., & Scala, A. (2020). Selective exposure shapes the Facebook news diet. PLoS ONE, 15(3), 1–17. https://doi.org/10.1371/journal.pone.0229129

Dafonte-Gómez, A. (2018). Audience as Medium: Motivations and Emotions in News Sharing. International Journal of Communication (19328036), 12, 2133–2152.

de Beer, D., & Matthee, M. (2020). Approaches to Identify Fake News: A Systematic Literature Review. Integrated Science in Digital Age 2020, 136, 13–22. https://doi.org/10.1007/978-3-030-49264-9_2

Department of Homeland Security. (2018). Countering False Information on Social Media in Disasters and Emergencies. https://www.dhs.gov/sites/default/files/publications/SMWG_Countering-False-Info-Social-Media-Disasters-Emergencies_Mar2018-508.pdf

Kumar, S. & Shah, N. (2018). False Information on Web and Social Media: A Survey. Social Media Analytics: Advances and Applications, by CRC press, 2018. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and Its Correction: Continued Influence and Successful Debiasing. Psychological Science in the Public Interest, 13(3), 106–131. https://doi.org/10.1177/1529100612451018

Maitland, N. & Lynch, J. (2020, March 1). SOCIAL MEDIA, ETHICS, AND THE PRIVACY PARADOX. Journal of Internet Law, 23(9), 3 - 14.

Newman, N., Fletcher, R., Levy, D., & Nielsen, R.K. (2016). Reuters Institute Digital News Report 2016, Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/research/files/Digital%2520News%2520Report%25202016.pdf

Pedersen, A., Griffiths, B., Watt, S. E. (2008). Attitudes toward out-groups and the perception of consensus: All feet do not wear one shoe. Journal of Community & Applied Social Psychology, 18, 543–557.

van der Linden, S., Leiserowitz, A., Rosenthal, S., and Maibach, E. (2017). Inoculating the public against misinformation about climate change. Global Challenges 1:1600008. doi: 10.1002/gch2.201600008

van der Linden, S., Roozenbeek, J., & Compton, J. (2020). Inoculating Against Fake News About COVID-19. Frontiers in Psychology, 11, N.PAG.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science 359, 6380 (2018), 1146–1151.

Westerman, D., Spence, P.R. and Van Der Heide, B. (2014), Social Media as Information Source: Recency of Updates and Credibility of Information. J Comput-Mediat Comm, 19: 171-183. doi:10.1111/jcc4.12041

Yu, F., Liu, Q., Wu, S., Wang, L., & Tan, T. (2019). Attention-based convolutional approach for misinformation identification from massive and noisy microblog posts. Computers & Security, 83, 106–121. https://doi.org/10.1016/j.cose.2019.02.003