Fighting Falsities – Addressing the Fake News Problem on Social Media

David Mednikov

Abstract—**Social Media has changed our society in many ways. A picture, video, or article can become viral within hours and make its way around the world instantly. The viral nature of social media allows propaganda and fake news to propagate across an individual's entire social network well before the truth has time to catch up.**

There will always be propaganda, but the nature of social media, where users are shown content that they are likely to click on or share, makes this even more dangerous. Fake news has affected millions around the world, and many of them are not even aware of it.

We envision a product that will be integrated with Facebook and Twitter (as well as other major social networks) to flag content as fake news or propaganda as soon as it is posted, rather than rely on users to report it.

I. The Problem I've Noticed

I know I am not the only one that has noticed this – the Internet is loaded with fake news all over the place, and no place spreads it faster than social media. Social media uses complex algorithms to show users content that their browsing habits indicate they would click/share.

This allows false stories to reach millions of people, with the intention of pushing a false narrative to disingenuously sway public opinion, well before the truth can reach them.

Often, the truth is lost among the chaos that fake news creates. This affects elections and governments. People voting when they don't have the facts is bad. People voting when they don't have the facts but *think* they do is worse.

We've all read about it in the news and experienced it firsthand. I see it when I make the mistake of going on Facebook; my parents (especially my father) consume it. You see a friend's parent post a wacky link on Facebook and you just think "how the hell…?"

Then you realize that 20 people liked it, and you begin to understand the magnitude of it.

It happens every day, to people of all ages, in all countries. It is a global problem. Bad actors are able to use fake news to mislead the public and to de-legitimize the press, which is meant to keep governments in check.

When the press is weakened, all opposition is weakened, and authoritarian governments are enabled. Turkey and Russia are two present-day examples of places where this occurred. We cannot allow this to continue to happen.

II. Evidence For This Problem

Brexit and the election of Donald Trump are two obvious examples of fake news affecting the outcome of an election.

The Washington Post wrote an article about how fake news helped Trump win here: https://www.washingtonpost.com/news/the-fix/wp/2018/04/03/a-new-study-suggests-fake-news-might-have-won-donald-trump-the-2016-election/

The Hill wrote a similar article about this as well: http://thehill.com/policy/cybersecurity/381449-researchers-say-fake-news-had-substantial-impact-on-20

I believe that there are three main reasons that fake news is so effective:

A. Wide-Reaching With Immediate Results

Content spreads on the Internet very quickly. Facebook and Twitter show users content that they are likely to click on, so as soon as fake news is posted by a page that a user follows they will see it and may even be notified of it.

These posts and advertisements are often targeted towards particular people, demographics, and/or personalities to make them more believable, which makes them more likely to view it and share it.

Fake news can reach hundreds of friends/followers in one's social network very quickly. Each person that shares it is introducing hundreds of new people to it. As it propagates across the entire spectrum of social media it only spreads faster.

See NBC's article about lies spreading faster than the truth here: https://www.nbcnews.com/health/health-news/fake-news-lies-spread-faster-social-media-truth-does-n854896

B. Confirmation Bias

The people that are most likely to believe fake news believe it because it lines up with their own beliefs. This is called confirmation bias, where an individual is likelier to believe something if they agree with the narrative behind it.

For example, if I believe that Hillary Clinton is a Satanist witch, and I see an article about her leading a pedophilia ring, I might be inclined to believe it. Also, the issue of groupthink arises: if 20 people like a post, that gives it legitimacy. As more people view or share a fake news post, it becomes more and more of a threat.

See USA Today's article about this: https://usatoday.com/story/money/columnist/2018/05/15/fake-news-social-media-

Submitted on Friday, June 29, 2018. Written in accordance with the course requirements of the Introduction to Software Engineering I course offered by

David Mednikov is with the Big Baller Brand. Like Lonzo Ball he attended UCLA and unlike Lonzo Ball he graduated from there.

confirmation-bias-echo-chambers/533857002/

C. *Profitability*

Fake news generates a lot of money. Advertising online is a huge business, so ad publishers make money, ads sold on fake news sites make money, clicking a link makes money, etc.

Fake news relies on individuals too. Fake news authors and propagandists are paid decent compensation since they often work on behalf of governments or wealthy organizations. As long as people need to eat they will be willing to set their ethics aside to get paid.

Despite that being done at a lower level, the higher levels of this industry (such as the company *Cambridge Analytica*) are dealing with many millions of dollars.

See this article BBC wrote on the profitability of fake news: https://www.bbc.com/news/av/business-38919403/how-do-fake-news-sites-make-money

AdPerfect, a service offering advertising solutions for newspapers and publishers, also wrote an article on this for the BBC: https://www.bbc.com/news/av/business-38919403/how-do-fake-news-sites-make-money

III. Fake News In Real-Time

Facebook knows your browsing habits. They show you a post that they think you'll click. It links to a story with a powerful image and a message that resonates with your belief.

You share it with your social network, hoping other people can see what you did and get the feeling you felt. Only there's one problem. The image is 4 years old and from something unrelated. The story is made up, and any reported statistics are just as fake.

You just shared fake news, and before someone could tell you it's fake, 5 people already viewed it. Some of them shared it to their social network. The lie spreads everywhere before the truth catches up to those that are willing to listen to it.

IV. How Software Will Address This problem

For this software system to work, it would need to be widely adopted by the two major social media platforms where content is shared: Facebook and Twitter.

The product would require an elaborate algorithm that would anonymously track thousands of social media accounts, comment sections, and other sources that are associated with fake news. Similar to Hamilton 68, it would analyze these for trends, common phrases, linked articles, etc.

What websites are being linked to? What hashtags and word combinations are being clicked and shared? What region are these posts originating from? What time of day are they peaking at? Are they using a proxy server?

In order to operate this service we will need to maintain a large database with entities for articles, topics, accounts, websites, hashtag trends, phrases, account aliases, users, etc.

We would use this database to keep track of all of the various bits of information mentioned above. By keeping track of data such as this, we can create a model for what looks like fake news and what doesn't.

By combining several sources into one algorithm, we can provide a more complete picture than Facebook and Twitter currently provide. When a user on Twitter or Facebook creates or shares a post that our system determines to match the fake news criteria as stored in our database, that post will be flagged immediately.

The aforementioned social media sites currently flag news that other users have reported to be fake, but that is not as effective as real-time flagging, considering how fast the lies can spread.

A user's news feed would look mostly the same, but posts that are found to be fake will be visibly flagged to the user.

There will be a warning that shows some details about the source of the fake news without revealing identifying information, so that the source can continue to be monitored without being compromised.

Users that share content that is fake will be notified and flagged as well. Facebook and Twitter will need to have some sort of policy to address accounts that knowingly or repeatedly share fake news.

V. What Makes It Effective

Instead of relying on users to flag news as fake, it will be analyzed and flagged using artificial intelligence. This way we can be one step ahead of the fake news providers, and we can begin to address the issue of the social media echo chamber, where false stories do not get flagged for several reasons.

Currently, if I share a fake story on Facebook and everyone that sees it has similar opinions to mine, it could easily pass as real, since no one in my network would recognize it as fake, or at least call me out for it. Using artificial intelligence can help alleviate this problem, so that the first person to see it after I post it already sees that its likely fake.

VI. Three features

1) *Clear Picture of Current Prevalent Trends in Fake News*
   Users should be able to get a clear picture of what popular trends/words/phrases are being tossed around by fake news sites and accounts, as well as what websites they are linking to and what articles/photos they are sharing.
2) *Immediate Analysis and Flagging*
   Newly-posted content should be analyzed in real-time so that our system will know if it is fake as soon as it is posted. That way the first user to see it will already be notified that it may be fake.
3) *Fake News Penalties*
   Facebook and Twitter will financially penalize verified users that are flagged for repeated violations of the fake news policy. By taking away the financial incentive, or adding a financial liability to the equation, the root of the problem is easier to address. Non-verified users will have their accounts suspended for repeated offenses. If financial penalties do not deter verified accounts, their accounts will be suspended as well.

VII. How This Will Help The World

The effectiveness of fake news on huge social networks such as Facebook and Twitter would decrease significantly if the fire could be put out before it was started.

This product would make it easier and faster to flag fake news before it gets spread around disguised as the truth. Users would be more aware of the information they're seeing and sharing, and purveyors of fake news would be punished.

The world is more peaceful when its leaders are not authoritarians that lie to their own people. Stopping their manipulation of public opinion through social media is a good step.

VIII. What Is Our Solution?

Our product will have multiple features that will help us decrease the influence and frequency of fake news on social media:

- We will monitor social media accounts that are known for sharing/creating fake news. Phrases, hashtags, images, linked websites, and other relevant data posted by these accounts will be added to our database.
- Whenever a user shares or creates a post on Facebook or Twitter, the content of their post will automatically be checked against the database in real time.
- If our system determines that the posted content matches the criteria for fake news in our database, a modal window will pop up notifying the user that the content they are attempting to share has been flagged as fake news. The user will still have the ability to post the article, and may also appeal the "decision" to website admins.
- If users do decide to post content that they were warned is fake news, this will count as a "point" against their account. Website admins can determine if bans or financial penalties are appropriate when a certain number of "points" is reached.
- Any submission that was flagged as fake news by our system and still posted by the user will appear on everyone else's timeline with a red shade background to indicate that it is fake news or at least untrustworthy.
- Sometimes users will share content that is brand new and does not yet match anything in our database. Our system will retroactively check posts against our database to determine if they were in fact fake.
- Users will be able to view a report on a post on their feed that has been marked as fake news to get a better idea of why it was flagged. The report will not compromise the account(s) that we are monitoring, however it will let users know if any wording, phrases, images, or URL was found to be fake news.
- Similar to current behavior, users will be able to anonymously mark a post on their timeline as fake news. However, as this is prone to abuse, there will need to be a team that reviews these submissions, with response time ideally limited to no more than a few hours.
- To protect certain sources such as satire websites, our system will have the ability to whitelist certain websites even though they are not publishing "real news". Websites to be added to the whitelist will need to go through the appeals process as mentioned earlier.
- In addition to banning users as mentioned earlier, website admins will be able to view a full profile for each user, including a user report, which will paint a holistic picture of that user's habits in regards to the integrity of the content they post, and whether or not they are aware of it.

IX. How Well Will It Work?

It will take some time to get an accurate answer as to whether or not our system is doing an adequate job in addressing the fake news problem on the Internet.

The purpose of this product is two-fold: to make users aware that the content that they themselves are posting is fake news, as well as making users aware that content that they are seeing on their feed is fake news. Our main goal is to make users more aware and more skeptical of the things they read and share. We will not be able to silence alternate views, but we can lessen their influence.

Ideally, after months of this platform being used, users that used to share a lot of fake news that aligned with their views will have decreased the frequency with which they post said fake news, whether that be the result of a ban or the user practicing more restraint when posting such content.

Statistically speaking, we want to see a significant decrease in the number of shared links to websites that are known distributors of fake news. For example, a tool like this should lead to a noticeable decrease in links to websites such as Breitbart.com and RT.com. While we cannot say for sure that it will, that is one of our main goals. Often Facebook and Twitter are just used as springboards to other platforms dedicated to sharing fake news, and this should help decrease that dramatically.

Similarly, we will see a decrease in the prevalence of hashtags or phrases that have been co-opted by fake news pushers. A great example of this is the so-called #WalkAway movement, which was found to be mainly Russian trolls posing as left-leaning voters "walking away" from the Democratic party. Hashtags like this are meant only to sow discord and divide, and our product will address them.

If these efforts are found to be effective, the truth can make its way back into the mainstream. Those that deny the truth will face the same harsh reality that the deniers of previous generations eventually faced.

This will without a doubt strengthen media and journalism around the world and set the quality of their work apart from the lazy and malicious fake news reporting. Similarly, authoritarians, dictators, and malicious pundits will see their influence dwindle as more people become aware that they are simply liars.

Ideally, this would also be a damaging blow to the Trump presidency and their malicious online presence.

X. Conclusion

In conclusion, this tool to fight back against fake news will not be a perfect tool. The battle of defeating fake news cannot be won by simply banning all sources of fake news on Facebook and Twitter, as users that consume fake news will simply move to a new platform that is (in their own words) less "biased against conservative views". This will make the echo chamber effect even stronger.

Instead, we seek to educate users and make them more aware and skeptical when sharing or posting content on social media networks. Many of these users grew up without the Internet, when content that was in print could be assumed to be true if it had the production quality of journalism. These days, a website that simply looks nice does not by any means have to be legitimate. We want these users to be more careful when they view, post, or share content on social media.

By making the user base of social media more skeptical about automatically sharing anything that aligns with the views, we place more emphasis on the truth. Placing emphasis on the truth weakens those that seek to do harm using lies. If people like Putin, Erdogan, Trump, etc., lose their ability to manipulate and create their own version of the truth, the world will be a much safer place.