

**SCIENTIFIC AND TECHNOLOGICAL APPROACHES
FOR DETECTING AND ADDRESSING FAKE NEWS**

Summary

Constantinos DASKALAKIS

It is hard to propose a formal definition of “fake news” without entering philosophical debates about what a “fact” is or what “truth” is. Postponing such debates, a working definition of “fake news” is the intentional misrepresentation of something that happened or the presentation of entirely constructed information with the goal of deceiving one’s audience. “Fake news” can also be expanded to include the orchestrated promotion of articles in such a way that it appears that their promotion was led by real people.

Fake news are spread to achieve several goals. One is financial gain. A fake news website that receives lots of visitors makes advertising revenue. Financial gain can also be made from the reaction to fake news, which may e.g. influence a stock price. Another major motivation for spreading fake news is pushing a political, religious, or other agenda. In this case, fake news spreading may be a paid service, and fake news campaigns may be engineered to influence public opinion, promote some ideology, incite fear, instigate a protest, influence an election, etc.

Social bots and fake social network accounts play a huge role in spreading fake news. Organized campaigns employ farms of social bots and fake accounts to like, share and retweet fake news in order to trick the algorithms of Facebook, Google, Twitter, etc., which determine what piece of information gets priority in news feeds and news searches, into perceiving a piece of fake news as socially engaging and worthy of promoting. Once the piece of fake news hits its intended audience, it is engineered to mislead its readers and incite intensive sharing, retweeting, and liking amplifying the benefits of social network effects.

Detecting fake news can be a difficult task even for trained human readers and fact checkers. Unfortunately, it is currently not technically possible to automatically detect “fake news” with decent accuracy say 90%. This determination is a complex task that pertains to the source of the news article and the history of what other articles this source has produced, its pattern of spreading, the presence of provocative or clickbait titles, the quality of its language, the sentiments it expresses, and other properties of the article itself, as well as the content of other articles on the same topic that have been published by reliable sources. Developing fake news detection technology is an active area of research that spans several fields including Algorithms, Artificial Intelligence, Natural Language Processing, Linguistics, Network Science, Psychology, etc.

While automating the process of fake news detection in its entirety is not possible with present technology, there are ways to use technology to aid fake news determination. For example, there exists technology for comparing the body of an article to its title, determining the quality of syntax and grammar in an article, detecting inflammatory language, performing sentiment analysis, etc. This technology can be used

to process the content of an article and help fake news detectors. There also exists technology that can analyse an article's spreading pattern. For example, network algorithms and statistical analyses can be used to identify motifs of social bot behaviour.

It is also important to attack the spreading of fake news by improving the technology that provides the very means via which fake news spreads around on the first place. Online platforms have already started introducing measures for curbing both the spreading of fake news and (relatedly) the profitability of fake news outlets. For example, Google has blacklisted abusive websites from participation in its online advertising platform, and is implementing changes into its search engine to prevent unreliable content from being ranked high. Facebook and Twitter have mechanisms for reporting, detecting and suspending abusive accounts. A powerful resource for content filtering is also asking users to flag and report fake news. Of course, this is quite tricky and should be implemented in an intelligent and robust manner so as to prevent users from flagging content with which they just disagree due political, religious or other beliefs. Wikipedia could offer inspiration for how such functionality could be implemented. Even merely flagging an article as controversial would be a good first cut.

The credibility of an article is also related to the credibility of its source, and there are mechanisms to differentiate between "reliable" news sources and "unreliable" ones. Admittedly there is no universal agreement on which news outlets are "reliable" as this is often entangled with the political leaning of the outlets and those of the one who is making the judgment. For example, in the US the New York Times is unreliable for many republicans, and Fox news is unreliable for many democrats. Leaving partisanship aside, these outlets have had a persistent presence for years and can certainly be deemed "somewhat reliable," at least until compelling evidence disqualifies them as such. On the other hand, fake news outlets often have a faster turn-over from being created to disappearing. A news outlet that has had a short life, and has received many red flags on its content should have a low credibility ranking. A news outlet that has had a short life, but has no red flags or just has a few of them should have higher credibility. If this news outlet has been cited positively by a "reliable" news outlet, this should positively boost its score, etc. Such a credibility rating would be similar to the use of star ratings in Yelp and other platforms, and is also used behind the scenes in search engines for ranking content.

I also find that fighting fake news can be dramatically helped by educating Internet users. Besides standard perception biases, the spreading of fake news exploits a lack of Internet literacy that should make one critical of information that they encounter on the Internet. As much as a reader may like a news piece to be factually true to support their ideology, they would be less likely to use it as evidence to support their ideas if they were vigilant about the potential inaccuracy of the information it contains. To offer an analogy, humans are trained to be critical to claims made in commercials and are less likely to use them as hard evidence to support their arguments. As humanity is adapting to the Internet era it is natural for humans to be out-of-sync with new technologies and the intricate ecosystems that they create around information production, consumption and circulation. Educating Internet users about how online advertising, search engines, and online social networks work as well as how these technologies can be manipulated will help combat fake news.

Last but not least, I would be strongly against a hard block of access to any website unless there is overwhelming evidence that it is a source of spam and abusive content. Given that detecting fake news is technologically challenging, and may be a hard task even for expert fact checkers and may even become subjective, I would not be willing to accept a hard censorship of any source of information unless there is overwhelming evidence of abuse. Freedom of communication and expression and the equal opportunity

to reach an audience for one's ideas must be protected. I would be more willing to entertain some form of soft block that may e.g. tag news sources as potential spam or of contentious content.