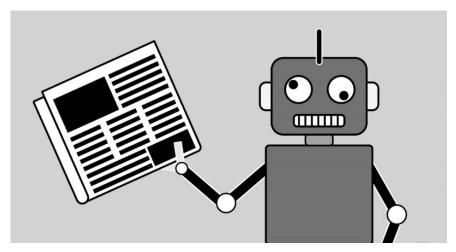# Even the Best AI for Spotting Fake News Is Still Terrible

It should be possible to automatically identify dubious news sources—but we'll need a lot more data

Still from an animation by Ms. Tech

By Karen Hao

When Facebook chief executive Mark Zuckerberg promised Congress that AI would help solve the problem of fake news, he revealed little in the way of *how*. New research brings us one step closer to figuring that out.

In an extensive study that will be presented at a conference later this month, researchers from MIT, Qatar Computing Research Institute (QCRI), and Sofia University in Bulgaria tested over 900 possible variables for predicting a media outlet's trustworthiness—probably the largest set ever proposed.

The researchers then trained a machine-learning model on different combinations of the variables to see which would produce the most accurate results. The best model accurately labeled news outlets with "low," "medium," or "high" factuality just 65% of the time.

This is far from a smashing success. But the experiments reveal important things about what it would take to outsource our fact-checking to a machine. Preslav Nakov, a senior scientist at QCRI and one of the researchers on the study, says he's optimistic that sources of fake news can automatically be spotted this way.

But that doesn't mean it will be easy.

Method to madness

In the explosion of research on fake-news detection since the 2016 US presidential campaign, four main approaches have emerged: fact-checking individual claims, detecting fake articles, hunting down trolls, and measuring the reliability of news sources. Nakov and the rest of the team chose to focus on the fourth because it gets closest to the origin of misinformation. It has also been studied the least.

Previous studies tried to characterize the reliability of a news source by how many of its claims matched or conflicted with claims that had been fact-checked already. In other words, a machine would compare the history of factual claims made by a news outlet against the conclusions of sites like Snopes or PolitiFact. The mechanism, however, relies on human fact-checking and evaluates the history of the outlet, not the immediate present. By the time the latest claims have been manually fact-checked, "it's already too late," says Nakov.

To spot a fake news source in close to real time, Nakov and his collaborators trained their system using variables that could be tabulated independently of human fact-checkers. These included analyses of the content, like the sentence structure of headlines and the word diversity in articles; overall site indicators, like the URL structure and website traffic; and measures of the outlet's influence, like its social-media engagement and Wikipedia page, if any.

To select the variables, the researchers relied both on previous research —past studies have shown that fake news articles tend to have

repetitive word choices, for example—and on new hypotheses.

By testing different combinations of variables, the researchers were able to identify the best predictors for a news source's reliability. Whether an outlet had a Wikipedia page, for example, had an outsize predictive power; the outlet's traffic, in contrast, had none. The exercise helped the researchers determine additional variables they could explore in the future.

Data starved

But there is one other obstacle: a shortage of training data—what Nakov calls the "ground truth."

For most machine-learning tasks, it's simple enough to annotate the training data. If you want to build a system that detects articles about sports, you can easily label articles as related or unrelated to that topic. You then feed the data set into a machine so it can learn the characteristics of a sports article.

But labeling media outlets with high or low factuality is much more sensitive. It must be done by professional journalists who follow rigorous methodologies, and it is a time-intensive process. As a result, it is challenging to build up a solid corpus of training data, which is partly why the accuracy of the study's model is so low. "The most obvious way to increase the accuracy is to get more training data," says Nakov

Currently, Media Bias Fact Check, the organization chosen to supply the "ground truth" for the research, has evaluated 2,500 media sources —a paucity in machine-learning terms. But Nakov says the organization's database is growing quickly. In addition to obtaining more training data, the researchers are also looking to improve their model's performance with more variables, some of which describe the structure of the website, whether it has contact information, and its patterns of publishing and deleting content.

They are also in the early stages of building a news aggregation platform that gives readers important cues to the trustworthiness of every story and source shared.

Despite the work left to be done, Nakov thinks such technology can help resolve the fake-news epidemic relatively quickly if platforms like

Facebook and Twitter earnestly exert the effort. "It is like fighting spam," he wrote in a Skype message. "We will never stop fake news completely, but we can put them under control."