

# The language gives it away: How an algorithm can help us detect fake news

August 15 2019, by Fatemeh Torabi Asr

---



Credit: AI-generated image ([disclaimer](#))

Have you ever read something online and shared it among your networks, only to find out it was false?

As a [software engineer](#) and computational linguist who spends most of her work and even leisure hours in front of a computer screen, I am

concerned about what I read online. In the age of social media, [many of us consume unreliable news sources](#). We're exposed to a wild flow of information in our social networks—especially if we spend a lot of time scanning our friends' random posts on Twitter and Facebook.

My colleagues and I at the [Discourse Processing Lab at Simon Fraser University](#) have conducted research on the linguistic characteristics of fake news.

## The effects of fake news

A study in the United Kingdom found that [about two-thirds of the adults surveyed regularly read news on Facebook, and that half of those had the experience of initially believing a fake news story](#). Another study, conducted by researchers at the Massachusetts Institute of Technology, focused on the cognitive aspects of exposure to fake news and [found that, on average, newsreaders believe a false news headline at least 20 percent of the time](#).

False stories are now [spreading 10 times faster than real news](#) and the problem of fake news seriously threatens our society.

For example, during the 2016 election in the United States, an astounding number of U.S. citizens believed and shared a patently false conspiracy claiming that Hilary Clinton was connected to [a human trafficking ring run out of a pizza restaurant](#). The owner of the restaurant received death threats, and one believer showed up in the restaurant with a gun. This —and [a number of other fake news stories](#) distributed during the election season —had an undeniable impact on people's votes.

It's often difficult to find the origin of a story after partisan groups, social media bots and friends of friends [have shared it thousands of times](#). Fact-checking websites such as [Snopes](#) and [Buzzfeed](#) can only

address a small portion of the most popular rumors.

The technology behind the internet and social media has enabled this spread of misinformation; maybe it's time to ask what this technology has to offer in addressing the problem.

## **Giveaways in writing style**

Recent advances in [machine learning](#) have made it possible for computers to instantaneously complete tasks that would have taken humans much longer. For example, there are computer programs that help police identify criminal faces [in a matter of seconds](#). This kind of artificial intelligence trains algorithms to classify, detect and make decisions.

When machine learning is applied to [natural language](#) processing, it is possible to build text classification systems that recognize one type of text from another.

During the past few years, natural language processing scientists have become more active in building algorithms to detect misinformation; this helps us to understand the characteristics of fake news and develop technology to help readers.

One approach finds relevant sources of information, assigns each source a credibility score and then integrates them to confirm or debunk a given claim. This approach is heavily dependent on [tracking down the original source of news and scoring its credibility based on a variety of factors](#).

A second approach examines the writing style of a news article rather than its origin. The linguistic characteristics of a written piece can tell us a lot about the authors and their motives. For example, specific words and phrases tend to [occur more frequently in a deceptive text compared](#)

[to one written honestly.](#)

## Spotting fake news

Our research identifies linguistic characteristics to detect fake news using machine learning and natural language processing technology. Our analysis of a large collection of fact-checked news articles on a variety of topics shows that, on average, fake news articles use more expressions that are common in hate speech, as well as words related to sex, death and anxiety. Genuine news, on the other hand, contains a larger proportion of words related to work (business) and money (economy).

This suggests that a stylistic approach combined with machine learning might be useful in detecting suspicious news.

Our fake news detector is built based on linguistic characteristics extracted from a large body of news articles. It takes a piece of text and shows how similar it is to the fake news and real news items that it has seen before. ([Try it out!](#))

The main challenge, however, is to build a system that can handle the vast variety of news topics and the quick change of headlines online, because computer algorithms learn from samples and if these samples are not sufficiently representative of online news, [the model's predictions would not be reliable.](#)

One option is to have human experts collect and label a large quantity of fake and real news articles. This data enables a machine-learning algorithm to find common features that keep occurring in each collection regardless of other varieties. Ultimately, the algorithm will be able to distinguish with confidence between previously unseen real or fake [news](#) articles.

**More information:** Discourse Processing Lab: [fake-news.research.sfu.ca/](https://fake-news.research.sfu.ca/)

Gordon Pennycook et al. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning, *Cognition* (2018). [DOI: 10.1016/j.cognition.2018.06.011](https://doi.org/10.1016/j.cognition.2018.06.011)

Martin Flintham et al. Falling for Fake News, *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18* (2018). [DOI: 10.1145/3173574.3173950](https://doi.org/10.1145/3173574.3173950)

Fatemeh Torabi Asr et al. Big Data and quality data for fake news and misinformation detection, *Big Data & Society* (2019). [DOI: 10.1177/2053951719843310](https://doi.org/10.1177/2053951719843310)

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: The language gives it away: How an algorithm can help us detect fake news (2019, August 15) retrieved 27 April 2024 from <https://techxplore.com/news/2019-08-language-algorithm-fake-news.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.