

April Fools hoax stories could offer clues to help identify 'fake news'

March 29 2019



Credit: CC0 Public Domain

Studying April Fools hoax news stories could offer clues to spotting 'fake news' articles, new research reveals.

Academic experts in Natural Language Processing from Lancaster University who are interested in deception have compared the language used within written April Fools hoaxes and [fake news](#) stories.

They have discovered that there are similarities in the written structure of humorous April Fools hoaxes—the spoof articles published by media outlets every April 1st—and malicious fake news stories.

The researchers have compiled a novel dataset, or corpus, of more than 500 April Fools articles sourced from more than 370 websites and written over 14 years.

"April Fools hoaxes are very useful because they provide us with a verifiable body of deceptive texts that give us an opportunity to find out about the linguistic techniques used when an author writes something fictitious disguised as a factual account," said Edward Dearden from Lancaster University, and lead-author of the research. "By looking at the language used in April Fools and comparing them with fake news stories we can get a better picture of the kinds of language used by authors of disinformation."

A comparison of April Fools hoax texts against genuine news articles written in the same period—but not published on April 1st—revealed stylistic differences.

Researchers focused on specific features within the texts, such as the amount of details used, vagueness, formality of writing style and complexity of language.

They then compared the April Fools stories with a 'fake news' dataset, previously compiled by a different team of researchers.

Although not all of the features found in April Fools hoaxes were found

to be useful for detecting fake news, there were a number of similar characteristics found across both.

They found April Fools hoaxes and fake news articles tend to contain less complex language, an easier reading difficulty, and longer sentences than genuine news.

Important details for news stories, such as names, places, dates and times, were found to be used less frequently within April Fools hoaxes and fake news. However, proper nouns, such as the names of prominent politicians 'Trump' or 'Hillary', are more abundant in fake news than in genuine news articles or April Fools, which have significantly fewer.

First person pronouns, such as 'we', are also a prominent feature for both April Fools and fake news. This goes against traditional thinking in deception detection, which suggests liars use fewer first person pronouns.

The researchers found that April fools hoax stories, when compared to genuine news:

- Are generally shorter in length
- Use more unique words
- Use longer sentences
- Are easier to read
- Refer to vague events in the future
- Contain more references to the present
- Are less interested in past events
- Contain fewer proper nouns
- Use more first person pronouns

Fake news stories, when compared to genuine news:

- Are shorter in length
- Are easier to read
- Use simplistic language
- Contain fewer punctuation marks
- Contain more proper nouns
- Are generally less formal—use more first names such as 'Hillary' and contain more profanity and spelling mistakes
- Contain very few dates
- Use more first person pronouns

The researchers also created a machine learning 'classifier' to identify if articles are April Fools hoaxes, fake news or genuine news stories. The classifier achieved a 75 per cent accuracy at identifying April Fools articles and 72 per cent for identifying fake news stories. When the classifier was trained on April Fools hoaxes and set the task of identifying fake news it recorded an accuracy of more than 65 per cent.

Dr. Alistair Baron, co-author of the paper, said: "Looking at details and complexities within a text are crucial when trying to determine if an [article](#) is a hoax. Although there are many differences, our results suggest that April Fools and fake news articles share some similar features, mostly involving structural complexity.

"Our findings suggest that there are certain features in common between different forms of disinformation and exploring these similarities may provide important insights for future research into deceptive news stories."

The research has been outlined in the paper 'Fool's Errand: Looking at April Fools Hoaxes as Disinformation through the Lens of Deception and Humour', which will be presented at the 20th International Conference on Computational Linguistics and Intelligent Text Processing, to be held in La Rochelle in April.

Provided by Lancaster University

Citation: April Fools hoax stories could offer clues to help identify 'fake news' (2019, March 29) retrieved 27 April 2024 from <https://phys.org/news/2019-03-april-hoax-stories-clues-fake.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.