

# Researchers develop early warning system to fight disinformation online

24 March 2020, by Jessica Sieff



Credit: CC0 Public Domain

Researchers at the University of Notre Dame are using artificial intelligence to develop an early warning system that will identify manipulated images, deepfake videos and disinformation online. The project is an effort to combat the rise of coordinated social media campaigns to incite violence, sow discord and threaten the integrity of democratic elections.

The scalable, automated system uses content-based image retrieval and applies [computer vision](#)-based techniques to root out political memes from multiple social networks.

"Memes are easy to create and even easier to share," said Tim Weninger, associate professor in the Department of Computer Science and Engineering at Notre Dame. "When it comes to political memes, these can be used to help get out the vote, but they can also be used to spread inaccurate information and cause harm."

Weninger, along with Walter Scheirer, an assistant professor in the Department of Computer Science and Engineering at Notre Dame, and members of

the research team collected more than two million images and content from various sources on Twitter and Instagram related to the 2019 general election in Indonesia. The results of that election, in which the left-leaning, centrist incumbent garnered a majority vote over the conservative, populist candidate, sparked a wave of violent protests that left eight people dead and hundreds injured. Their [study](#) found both spontaneous and coordinated campaigns with the intent to influence the election and incite violence.

Those campaigns consisted of manipulated images exhibiting false claims and misrepresentation of incidents, logos belonging to legitimate [news sources](#) being used on fabricated news stories and memes created with the intent to provoke citizens and supporters of both parties.

While the ramifications of such campaigns were evident in the case of the Indonesian general election, the threat to democratic elections in the West already exists. The research team at Notre Dame, comprised of digital forensics experts and specialists in peace studies, said they are developing the system to flag manipulated content to prevent violence, and to warn journalists or [election](#) monitors of potential threats in real time.

The system, which is in the research and development phase, would be scalable to provide users with tailored options for monitoring content. While many challenges remain, such as determining an optimal means of scaling up data ingestion and processing for quick turnaround, Scheirer said the system is currently being evaluated for transition to operational use.

Development is not too far behind when it comes to the possibility of monitoring the 2020 [general election](#) in the United States, he said, and their team is already collecting relevant data.

"The disinformation age is here," said Scheirer. "A

deepfake replacing actors in a popular film might seem fun and lighthearted but imagine a video or a meme created for the sole purpose of pitting one world leader against another—saying words they didn't actually say. Imagine how quickly that content could be shared and spread across platforms. Consider the consequences of those actions."

**More information:** Michael Yankoski et al, An AI early warning system to monitor online disinformation, stop violence, and protect elections, *Bulletin of the Atomic Scientists* (2020). DOI: [10.1080/00963402.2020.1728976](https://doi.org/10.1080/00963402.2020.1728976)

Provided by University of Notre Dame

APA citation: Researchers develop early warning system to fight disinformation online (2020, March 24) retrieved 9 April 2020 from <https://techxplore.com/news/2020-03-early-disinformation-online.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.*