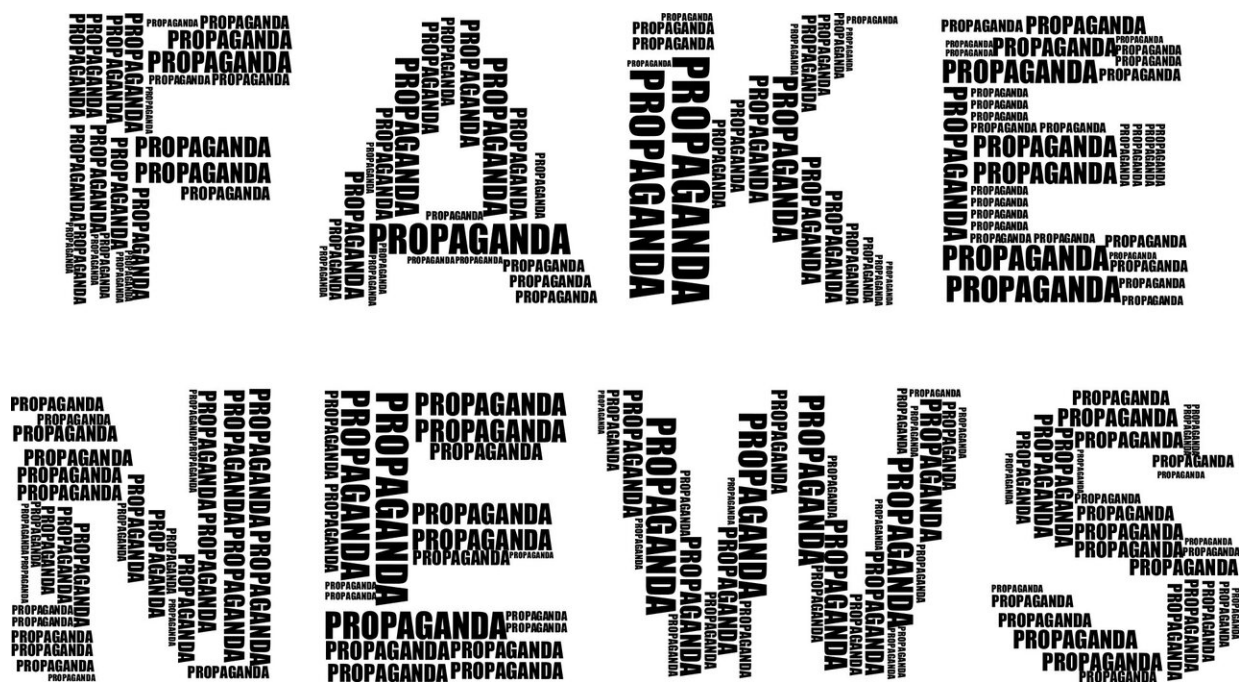


# How technology can detect fake news in videos

June 29 2022, by Agustín López



Credit: Pixabay/CC0 Public Domain

Social media represent a major channel for the spreading of fake news and disinformation. This situation has been made worse with recent advances in photo and video editing and artificial intelligence tools, which make it easy to tamper with audiovisual files, for example with so-called deepfakes, which combine and superimpose images, audio and video clips to create montages that look like real footage.

Researchers from the K-riptography and Information Security for Open Networks (KISON) and the Communication Networks & Social Change (CNSC) groups of the Internet Interdisciplinary Institute (IN3) at the Universitat Oberta de Catalunya (UOC) have launched a new project to develop innovative technology that, using artificial intelligence and data concealment techniques, should help users to automatically differentiate between original and adulterated [multimedia content](#), thus contributing to minimizing the reposting of fake news. DISSIMILAR is an international initiative headed by the UOC including researchers from the Warsaw University of Technology (Poland) and Okayama University (Japan).

"The project has two objectives: firstly, to provide [content creators](#) with tools to watermark their creations, thus making any modification easily detectable; and secondly, to offer [social media](#) users tools based on latest-generation signal processing and machine learning methods to detect fake digital content," explained Professor David Megías, KISON lead researcher and director of the IN3. Furthermore, DISSIMILAR aims to include "the cultural dimension and the viewpoint of the end user throughout the entire project," from the designing of the tools to the study of usability in the different stages.

## **The danger of biases**

Currently, there are basically two types of tools to detect fake news. Firstly, there are automatic ones based on machine learning, of which (currently) only a few prototypes are in existence. And, secondly, there are the fake news detection platforms featuring human involvement, as is the case with Facebook and Twitter, which require the participation of people to ascertain whether specific content is genuine or fake.

According to David Megías, this centralized solution could be affected by "different biases" and encourage censorship. "We believe that an objective assessment based on technological tools might be a better

option, provided that users have the last word on deciding, on the basis of a pre-evaluation, whether they can trust certain content or not," he explained.

For Megías, there is no "single silver bullet" that can detect fake news: rather, detection needs to be carried out with a combination of different tools. "That's why we've opted to explore the concealment of information (watermarks), digital content forensics analysis techniques (to a great extent based on signal processing) and, it goes without saying, machine learning," he noted.

## **Automatically verifying multimedia files**

Digital watermarking comprises a series of techniques in the field of data concealment that embed imperceptible information in the original file to be able "easily and automatically" verify a multimedia file. "It can be used to indicate a content's legitimacy by, for example, confirming that a video or photo has been distributed by an official news agency, and can also be used as an authentication mark, which would be deleted in the case of modification of the content, or to trace the origin of the data. In other words, it can tell if the source of the information (e.g. a Twitter account) is spreading fake content," explained Megías.

## **Digital content forensics analysis techniques**

The project will combine the development of watermarks with the application of digital content forensics analysis techniques. The goal is to leverage signal processing technology to detect the intrinsic distortions produced by the devices and programs used when creating or modifying any audiovisual file. These processes give rise to a range of alterations, such as sensor noise or optical distortion, which could be detected by means of machine learning models. "The idea is that the combination of

all these tools improves outcomes when compared with the use of single solutions," stated Megías.

## **Studies with users in Catalonia, Poland and Japan**

One of the key characteristics of DISSIMILAR is its "holistic" approach and its gathering of the "perceptions and cultural components around fake news." With this in mind, different user-focused studies will be carried out, broken down into different stages. "Firstly, we want to find out how users interact with the news, what interests them, what media they consume, depending upon their interests, what they use as their basis to identify certain content as fake news and what they are prepared to do to check its truthfulness. If we can identify these things, it will make it easier for the technological tools we design to help prevent the propagation of fake news," explained Megías.

These perceptions will be gaged in different places and cultural contexts, in user group studies in Catalonia, Poland and Japan, so as to incorporate their idiosyncrasies when designing the solutions. "This is important because, for example, each country has governments and/or public authorities with greater or lesser degrees of credibility. This has an impact on how news is followed and support for [fake news](#): if I don't believe in the word of the authorities, why should I pay any attention to the [news](#) coming from these sources? This could be seen during the COVID-19 crisis: in countries in which there was less trust in the public authorities, there was less respect for suggestions and rules on the handling of the pandemic and vaccination," said Andrea Rosales, a CNSC researcher.

## **A product that is easy to use and understand**

In stage two, users will participate in designing the tool to "ensure that

the product will be well-received, easy to use and understandable," said Andrea Rosales. "We'd like them to be involved with us throughout the entire process until the final prototype is produced, as this will help us to provide a better response to their needs and priorities and do what other solutions haven't been able to," added David Megías.

This user acceptance could in the future be a factor that leads social network platforms to include the solutions developed in this project. "If our experiments bear fruit, it would be great if they integrated these technologies. For the time being, we'd be happy with a working prototype and a proof of concept that could encourage social media platforms to include these technologies in the future," concluded David Megías.

Previous research was published in the *Special Issue on the ARES-Workshops 2021*.

**More information:** D. Megías et al, Architecture of a fake news detection system combining digital watermarking, signal processing, and machine learning, *Special Issue on the ARES-Workshops 2021* (2022). [DOI: 10.22667/JOWUA.2022.03.31.033](https://doi.org/10.22667/JOWUA.2022.03.31.033)

A. Qureshi et al, Detecting Deepfake Videos using Digital Watermarking, *2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)* (2021). [ieeexplore.ieee.org/document/9689555](https://ieeexplore.ieee.org/document/9689555)

David Megías et al, DISSIMILAR: Towards fake news detection using information hiding, signal processing and machine learning, *16th International Conference on Availability, Reliability and Security (ARES 2021)* (2021). [doi.org/10.1145/3465481.3470088](https://doi.org/10.1145/3465481.3470088)

Provided by Universitat Oberta de Catalunya (UOC)

Citation: How technology can detect fake news in videos (2022, June 29) retrieved 27 April 2024 from <https://techxplore.com/news/2022-06-technology-fake-news-videos.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.