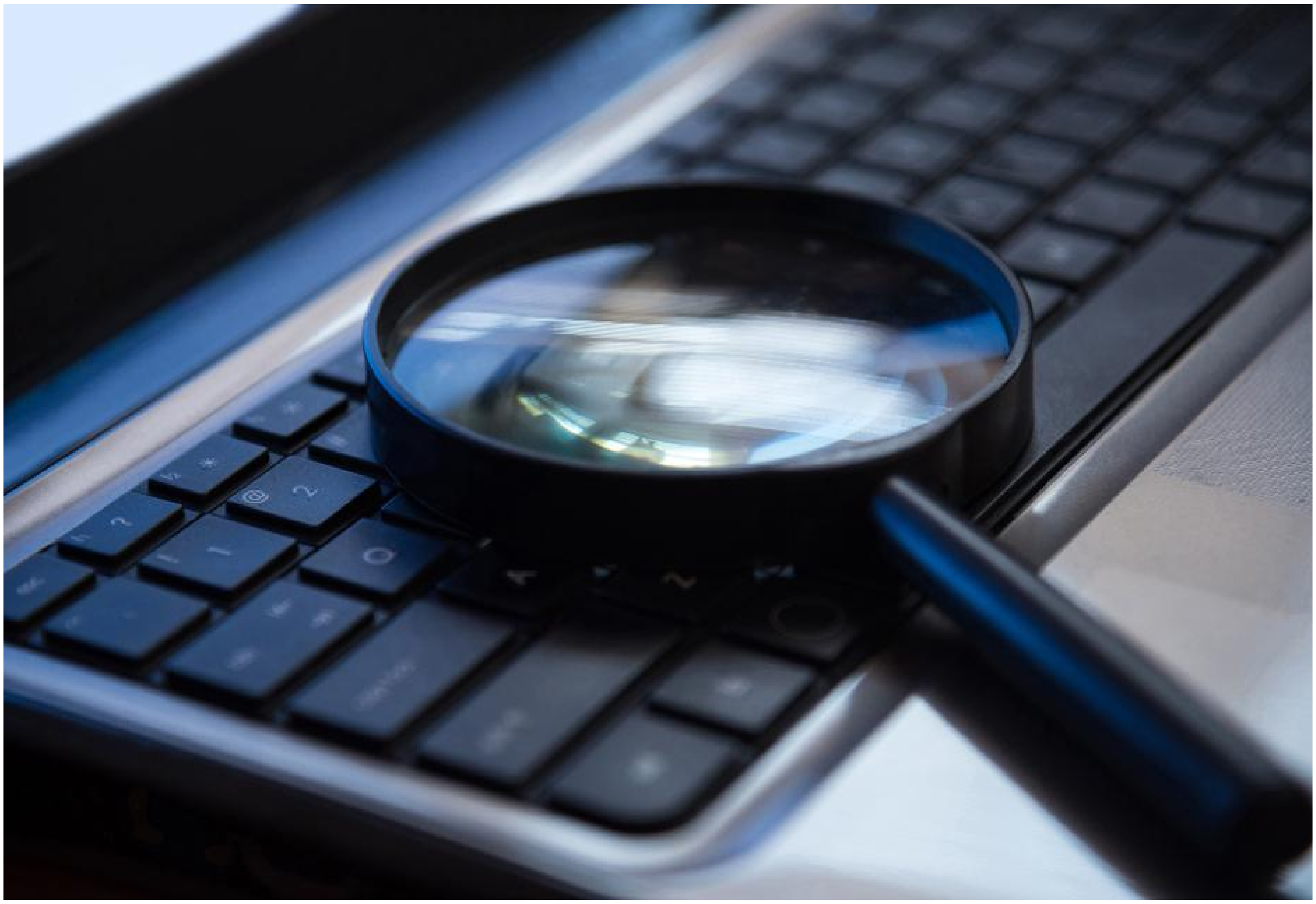
498 views | Oct 15, 2019, 02:13pm

# Unleashing New Weapons In The War On Fake News

**Charles Towers-Clark** Contributor ⓘ
AI & Big Data
*I write about AI, data, deep tech & self-management in the digital age*



Detecting fake news is getting more difficult as more false information pours onto the internet every day, and from very influential sources. New papers from MIT explore how current methods are failing, and bring new weapons to tl fight against fake FREEPIK

Now is the time to make facts great again. “Fake news,” the 2017 Collins word of the year, poses a serious threat to the values of honesty, truth, and accountability—values that purveyors of falsified information don’t seem to hold too closely. Apart from the most obvious dangers of spreading false information (erosion of trust, political or national hostility, widespread uncertainty) the prevalence of AI systems on social medi mean that unverified claims and slanderous falsehoods are picked up and distributed a eye-watering speeds.

The surge in fake news over recent years has wrought unprecedented effects on the national and international stage, and there is a heated debate between the public, governments, the media and big tech regarding how best to handle this wave of misinformation. But as is generally the case with AI, its ability to perpetuate and suppo fake news is evenly matched with its ability to fight it. New advances in identifying fals information and detecting machine-generated text using AI will help to curb the sprea of false information and cut off the more ludicrous claims at the source—provided that those in positions of influence are prepared to fight it.

### Fake-your-own Trump card

While the term fake news has been bandied around a lot since around 2016, the spread of falsified information is by no means diminishing—in fact the rise of deepfakes has added a far more sinister dimension to the war on fake news. Furthermore, those in positions of influence are doing little to stop fake news, and there is a lot of evidence to suggest that governments are as much to blame as individual actors. The most recent scandal to break in the age of fake news is once again linked to the Trump administration, relating to an allegedly false campaign ad about Joe Biden that was distributed on Facebook and Twitter - and which the social media giants refuse to remove. The advertisement, claiming that Biden coerced Ukraine to fire a prosecutor targeting his son Hunter has been refused by some news outlets due to inaccuracy, but arguably today's most influential distributor of news, Twitter and Facebook, have invoked their policies to support their decisions to continue hosting malevolent false claims.

Nick Clegg, the communications chief for Facebook, has said that the company will not verify the claims of politicians for factual accuracy, and Facebook's head of global elections policy, Katie Harbath, said in a letter to the Biden campaign that claims made by politicians are "considered direct speech and ineligible for our third-party fact checking program." Twitter has been less direct with its response to the matter, simply stating that the ad "is not in violation of [their] policies." Apart from the seemingly counterintuitive logic that political actors are immune to having their claims held accountable, the passive intervention by social media giants into political matters is disturbing (especially when there is financial interest involved). With the level of influence that social media holds over our daily lives, and the aggregation, recommendation, and trending algorithms that help to spread news faster than has eve been possible, a point-blank refusal to check political claims is a significant setback in the war on fake news.

Today In: Innovation

## Weapons of mass verification

There are more battles to come, however, and Artificial Intelligence will be a valuable tool in the fight against fake news—even if those with the most advanced AI would rath use it to proliferate unverified claims. Two new papers released today (October 15) out MIT CSAIL are using AI to shed light on this issue on two fronts. The first paper focuse on identifying text that has been generated by a machine, as this is often a source of fal information and a common vehicle for its rapid proliferation on social media. The second addresses issues in current fact-verification methods that rely on the FEVER dataset (the largest dataset for Fact Extraction and Verification in text) by using a different approach to verify claims and showing how bias can thwart verification algorithms.

These two papers tackle fake news in different ways, looking at how to limit the creatio of false information in the first place and how to identify claims more effectively withou relying on out of date information or biased data. “There’s a growing concern about machine-generated fake text, and for a good reason,” says MIT CSAIL PhD student Tal Schuster, lead author of the verification paper. These concerns, likely stemming from t use of bots to disseminate misleading information in recent years, are valid, but the correlation of fake news with machine-generated text is also problematic, as Schuster notes: “text generators don’t have a specific agenda - it’s up to the user to decide how t use this technology.” The team compared legitimate and false auto-completed text samples, and then used standard provenance-based detection methods to check the accuracy of each. In one example, an AI-generated article that accurately described findings by NASA scientists was deemed to be false, just because the article was generated by a machine. “We need to have the mindset that the most intrinsic ‘fake news' characteristic is factual falseness, not whether or not the text was generated by machines," says Schuster, and aside from this finding, the paper also discussed "the types of benchmarks that should be used to evaluate neural fake news detectors," to more accurately identifying fake news at the source.

The second paper looked at how algorithms trained on FEVER suffered significant bias due to overly simplistic reasoning. Models trained on FEVER check for accuracy agains Wikipedia articles, which in itself can be seen in two ways—as a self-policed open-sourc system utopia, or an infinitely corruptible source. One outcome of using this training dataset (as with any dataset that is not completely trustworthy), the team found, is that negative phrases such as "did not" or "yet to" are often classified as false, when this is n necessarily the case. These models focus on the language of the claim, and do not look a context or external evidence to ascertain validity. A further problem of classifying information without evidence is that an out of date claim (for instance, "Olivia Colman has never won an Oscar") could enter reasoning, despite being easily verifiable elsewhe (her IMDB profile). To tackle this, the team created a new dataset without such biases and created a model that improved over time through positive and negative reinforcement. "True claims with the phrase 'did not' would be upweighted, so that in the newly weighted dataset, that phrase would no longer be correlated with the 'false' class," says paper author Darsh J Shah, allowing incorrect classifications to be rectified over time.

### Finding & fighting fake news

These papers by MIT CSAIL show that current verification and detection systems are n sufficient to stop the swathes of false information that are currently plaguing media of a kinds. Whether it is a case of looking at the wrong identifiers of fake news, such as the source rather than the content, or verifying claims using biased, out of date, or un-evidenced data, our powers to stop fake news need to be upgraded.

With world leaders and the most influential tech companies actively and passively spreading fake news, it is crucial that our tools to detect and refute fake news are up to scratch. As we move into possibly the most factually turbulent period in history, we nee more investigation into just how we currently fight back against fake news, to help preserve the trust we have in each other, and the respect we hold for those that wield such control over our lives.

*Follow me on Twitter.*

**Charles Towers-Clark**

I have been working in the M2M, IoT, and data space since founding Pod Group (a provider of IoT connectivity & billing software) in 1999, and have become greatly int... **Read More**