# How artificial intelligence can detect - and create - fake news

By Anjana Susarla
4 May 2018

When Mark Zuckerberg told Congress Facebook would use artificial intelligence to detect fake news posted on the social media site, he wasn't particularly specific about what that meant. Given my own work using image and video analytics, I suggest the company should be careful. Despite some basic potential flaws, AI can be a useful tool for spotting online propaganda - but it can also be startlingly good at creating misleading material.
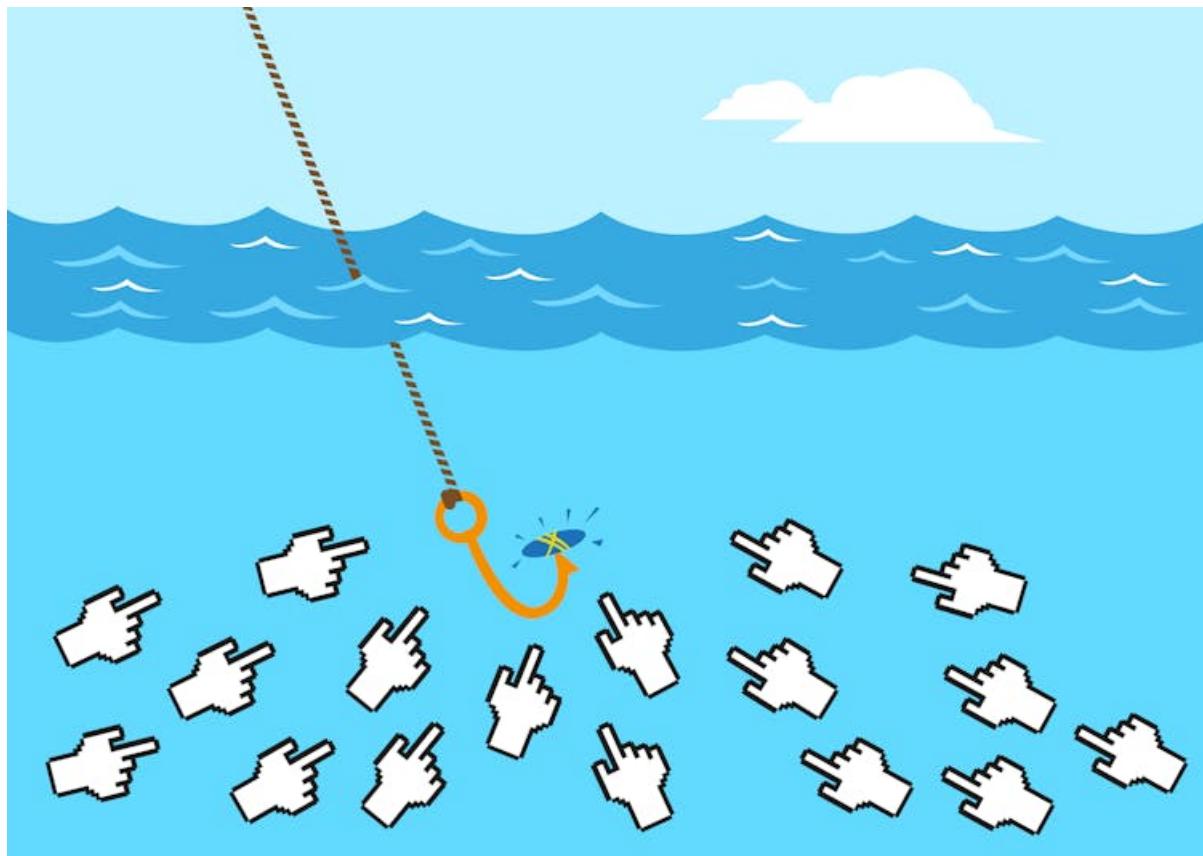
Is that clickbait true? Crystal Eye Studio/Shutterstock.com

This sure looks like Barack Obama saying some things he probably would never say.

Researchers already know that online fake news spreads [much more quickly and more widely](#) than real news. My research has similarly found that online posts with [fake medical information get more views](#), comments and likes than those with accurate medical content. In an online world where viewers have limited attention and are saturated with content choices, it often appears as though fake information is more appealing or engaging to viewers.

The problem is getting worse: By 2022, people in developed economies could be encountering [more fake news than real information](#). This could bring about a phenomenon researchers have dubbed "[reality vertigo](#)" – in which computers can generate such convincing content that regular people may have a hard time figuring out what's true anymore.

Detecting falsehood

Machine learning algorithms, one type of AI, have been successful for decades fighting spam email, by [analyzing messages' text](#) and determining how likely it is that a particular message is a real communication from an actual person – or a mass-distributed [solicitation for pharmaceuticals](#) or [claim of a long-lost fortune](#).

Building on this type of text analysis in spam-fighting, AI systems can evaluate how well a post's text, or a headline, [compares with the actual content](#) of an article someone is sharing online. Another method could examine similar articles to see whether other news media [have differing facts](#). Similar systems can [identify specific accounts and source websites](#) that spread fake news.

An endless cycle

However, those methods assume the people who spread fake news don't change their approaches. They often shift tactics, [manipulating the content of fake posts](#) in efforts to make them look more authentic.

Using AI to evaluate information can also [expose – and amplify – certain biases](#) in society. This can relate to gender, [racial background](#) or [neighbourhood stereotypes](#). It can even have political consequences, potentially [restricting expression of particular viewpoints](#). For example, YouTube has [cut off advertising](#) from certain types of video channels, costing their creators money.

Context is also key. Words' meanings can change over time. And the same word can mean different things on liberal sites and conservative ones. For example, a post with the terms "WikiLeaks" and "DNC" on a more liberal site could be more likely to be news, while on a conservative site it could refer to a [particular set of conspiracy theories](#).

Using AI to make fake news

The biggest challenge, however, of using AI to detect fake news is that it puts technology in [an arms race with itself](#). Machine learning systems are already proving spookily capable of creating what is called "deepfakes" – photos and videos that realistically replace one person's face with another, to make it appear that, for example, a celebrity was [photographed in a revealing pose](#) or a public figure is [saying things he'd never actually say](#). Even [smartphone apps](#) are capable of this sort of substitution – which makes this technology available to just about anyone, even without Hollywood-level video editing skills.

Researchers are already preparing to use AI to identify these AI-created fakes. For example, techniques for video magnification can [detect changes in human pulse](#) that would establish whether a person in a video is real or computer-generated. But both [fakers and fake-detectors will get better](#). Some fakes could become so sophisticated that they become very hard to rebut or dismiss – unlike [earlier generations of fakes](#), which used simple language and made easily refuted claims.

Human intelligence is the real key

The best way to combat the spread of fake news may be to depend on people. The societal consequences of fake news – greater political polarisation, increased partisanship, and eroded trust in mainstream media and government – are significant. If more people knew the stakes were that high, they might be more wary of information, particularly if it is more emotionally based, because that's an effective way to get people's attention.

When someone sees an enraging post, that person would do better to investigate the information, rather than sharing it immediately. The act of sharing also lends credibility to a post: When other people see it, they register that it was shared by someone they know and presumably trust at least a bit, and are less likely to notice whether the original source is questionable.

Social media sites like YouTube and Facebook could voluntarily decide to label their content, showing clearly whether an item purporting to be news is verified by a reputable source. Zuckerberg told Congress he wants to mobilise the "community" of Facebook users to direct his company's algorithms. Facebook could crowd-source verification efforts. Wikipedia also offers a model, of dedicated volunteers who track and verify information.

Facebook could use its partnerships with news organisations and volunteers to train AI, continually tweaking the system to respond to propagandists' changes in topics and tactics. This won't catch every piece of news posted online, but it would make it easier for large numbers of people to tell fact from fake. That could reduce the chances that fictional and misleading stories would become popular online.

Reassuringly, people who have some exposure to accurate news are better at distinguishing between real and fake information. The key is to make sure that at least some of what people see online is, in fact, true.

This article was originally published on The Conversation. Read the original article.</>

ABOUT THE AUTHOR

Anjana Susarla, Associate Professor of Information Systems, *Michigan State University*