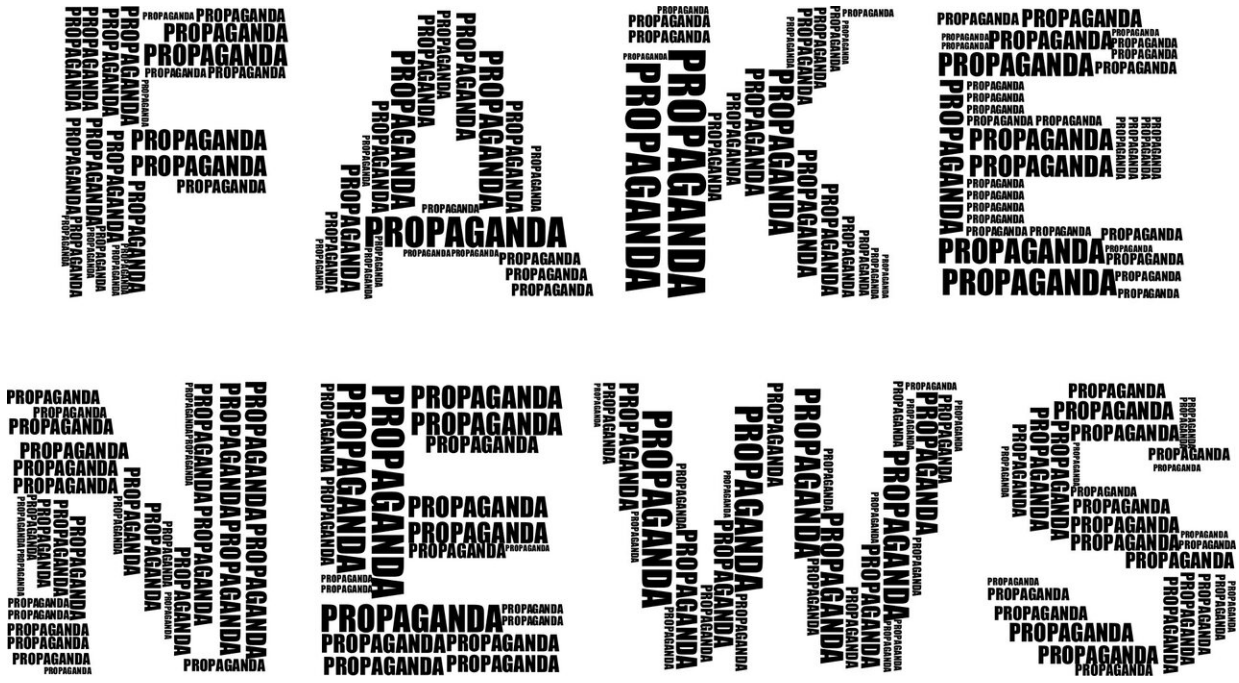
# Q&A: Artificial intelligence—stemming the tide of fake facts

November 21 2023




Credit: Pixabay/CC0 Public Domain

Professor Stefan Feuerriegel is Head of the Institute of Artificial Intelligence (AI) in Management at LMU, and his research focuses on the challenges of the digitalization wave.

In a commentary recently published in the journal *Nature Human Behavior*, he points out the dangers posed by AI-generated

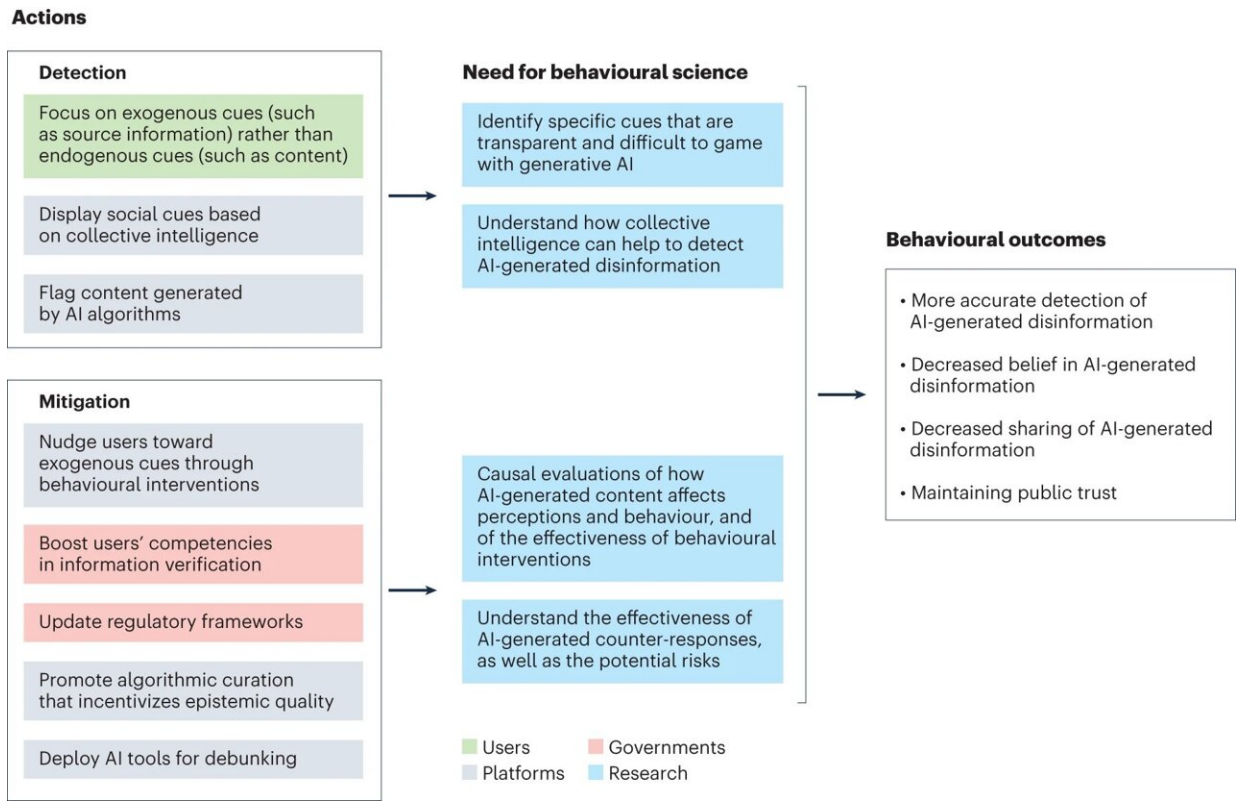
disinformation and suggests ways of countering them.

## Pretty much at the push of a button, AI tools are now able to generate convincing texts, images, voices, and even videos. Can we still trust our eyes, our ears, and our common sense going forward?

Stefan Feuerriegel: The picture of Pope Francis that went viral this past spring illustrates just how convincing such artificially generated content is. Ordinary people were scarcely equipped to recognize whether the picture was real or not. Of course, gifted artists were able to hoodwink the public and experts with counterfeit paintings in bygone days.

But in the case of the picture of the pope, it was one person who quickly spun out dozens of versions of this photo. This highlighted what's possible with AI today—producing these images is easy, and you don't need special skills.

At the moment, AI programs still don't get some individual details quite right: sometimes, the background is a bit off, and AI has a habit of rendering hands with too many or too few fingers. However, engineers are developing and optimizing the technology at a rapid rate. Over the next few months, the improvements will be apparent.




Role of behavioral science in the era of AI-generated disinformation. Behavioral science can promote detection and mitigation strategies for humans in tackling AI-generated disinformation. Credit: *Nature Human Behaviour* (2023). DOI: 10.1038/s41562-023-01726-2

## What makes AI-generated fake news particularly dangerous?

Unfortunately, not only are fakes with error-free text and authentic-looking pictures already very convincing, but AI makes it possible to personalize misinformation and tailor it to the religion, gender, or political convictions of individual consumers and sow rage or hatred in each target group.

You can build a bot that no longer just posts a message but writes to

people personally on Facebook or Twitter, responds to replies, and carries on conversations. This extends even to fake calls, where scammers or other bad actors deliberately generate a voice that sounds like a person's family member.

## When looking for typos and counting fingers is no longer enough, what are the telltale signs that can alert us to fakes?

As regards the content, there aren't any signs. Or at least there soon won't be. Either you rely on certain trusted sources, or you have to do your own research. That being said, you should be aware that purveyors of fake facts can obtain prominent search engine rankings with AI-generated websites and sources. In social networks, moreover, the images are so small that people can't even spot any mistakes that may be there.

There are low-resolution pictures and videos circulating on the internet that don't allow people to properly assess whether they are real or fake. This makes it difficult to distinguish misinformation from genuine content. In armed conflicts, the power of images and social media plays an important role and can have an explosive political impact. Here, too, we're already seeing the deployment of AI-generated material alongside conventional fakes.

## Are we entering a new age of disinformation?

Many of my colleagues are saying that we're actually already living in a fake news era. The problem is only going to get worse. The decisive question is: Who is exploiting these new possibilities? I'm less concerned about private individuals spreading misinformation through ChatGPT or DALL·E. Such people rarely have the range, or indeed the desire, to

exercise major influence. Instead, we must keep an eye on the actors that use these tools for large-scale disinformation campaigns, deliberately calibrate them, or even develop their own tools without any inbuilt security mechanisms.

Things get really dangerous when the big players enter the fray—say, a state actor in a non-democratic country with a certain political agenda. In the course of the Russian invasion of Ukraine, we've seen how much effort has been pumped into pro-Russian propaganda. Indeed, why wouldn't the responsible agencies work with these new tools, which enable them to respond faster and produce authentic-seeming content on a much larger scale? It would be naïve to believe that these opportunities would go unexploited.

## Are the fact checkers able to keep up with all this?

We know that human fact-checkers need up to 24 hours to verify a news story. By this time, it can long have gone viral. In current crises in particular, debunking misinformation before it spreads is a huge challenge. Facebook and Twitter are currently using generative AI to automatically identify fake news.

There are also discussions underway about employing watermarks, by means of which platforms could recognize and filter out AI-generated content. This requires the cooperation of platforms. It's no use if Twitter utilizes this tool, but the misinformation then goes viral on WhatsApp. Moreover, there can still be actors who do not respect the rules and program their own AI without watermarks.

## How can we counter the flood of AI misinformation?

Personally, I've become much more circumspect in the content I

consume. We must be a lot more alert as to what we read, especially on social networks. Unfortunately, many people are not aware of the problem. Although they know that AI is capable of producing high-quality texts and realistic pictures, it's not yet on their radar how this technology can be misused.

First of all, the platforms must be made to fulfill their responsibilities. They know the sources of the information. Users just see the posts, whereas the platforms can see whether some computer program is posting these contents in the background or a real person is behind them. Furthermore, the operators of the networks can remind their members to critically evaluate and check information. Platforms could also do a lot more to filter out fake news—some are making an effort in this regard, others not so much.

Secondly, we must inform and prepare every individual and the public at large. We need training courses on media skills and digital literacy that cover AI disinformation and are continuously updated as technology changes.

And then, thirdly, there is the question as to what politicians should do and how much regulation is useful and effective. This is a prickly topic because such measures can rub up against freedom of speech. AI is currently at the top of the agenda for the European Parliament, and I believe we could devise good solutions relatively swiftly in terms of regulatory frameworks.

## Are we ready for what's coming? Where do we need more research?

No, we're not adequately prepared. We're dealing with a new technology that we need to understand better, and that requires much more basic

research. Fortunately, a lot of research is being carried out in this area, not least at LMU. Linguists, sociologists, political scientists, and researchers from many other disciplines are probing this complex subject. Behavioral scientists are working on understanding how people react to such artificially generated information in the first place.

Legal scholars are studying the legal hurdles and seeking to balance the precious commodity of freedom of speech against solution-oriented, implementable approaches. Meanwhile, IT is tasked with finding out what is technologically feasible. We've got an interdisciplinary environment at LMU, where many different research fields are collectively developing a research agenda.

**More information:** Stefan Feuerriegel et al, Research can help to tackle AI-generated disinformation, *Nature Human Behaviour* (2023). DOI: 10.1038/s41562-023-01726-2

Provided by Ludwig Maximilian University of Munich