HOCHE 2023-2024 PC Mrs Mélières

oral sessions with C. Hamard

**www.theguardian.com**

**ChatGPT is making up fake Guardian articles. Here’s how we’re responding | Chris Moran**

Last month one of our journalists received an interesting email. A researcher had come across mention of a Guardian article, written by the journalist on a specific subject from a few years before. But the piece was proving elusive on our website and in search. Had it been removed intentionally from the website because of a problem we’d identified? The reporter couldn’t remember writing the specific piece, but the headline certainly sounded like something they would have written. It was a subject they were identified with and had a record of covering. They could find no trace of its existence. Why? Because it had never been written.

Luckily the researcher had told us that they had carried out their research using ChatGPT. In response to being asked about articles on this subject, the AI had simply made some up. Its fluency, and the vast training data it is built on, meant that the existence of the invented piece even seemed believable to the person who absolutely hadn’t written it.

Huge amounts have been written about generative AI’s tendency to manufacture facts and events. But this specific wrinkle – the invention of sources – is particularly troubling for trusted news organisations and journalists whose inclusion adds legitimacy and weight to a persuasively written fantasy. And for readers and the wider information ecosystem, it opens up whole new questions about whether citations can be trusted in any way.

If this seems like an edge case, it’s important to note that ChatGPT, from a cold start in November, registered 100 million monthly users in January. TikTok, unquestionably a digital phenomenon, took nine months to hit the same level. A recent study of 1,000 students in the US found that 89% have used ChatGPT to help with a homework assignment.

It’s easy to get sucked into the detail on generative AI, because it is inherently opaque. The ideas and implications, already explored by academics across multiple disciplines, are hugely complex, the technology is developing rapidly, and companies with huge existing market shares are integrating it as fast as they can to gain competitive advantages, disrupt each other and above all satisfy shareholders.

But the question for responsible news organisations is simple, and urgent: what can this technology do right now, and how can it benefit responsible reporting at a time when the wider information ecosystem is already under pressure from misinformation, polarisation and bad actors.

This is the question we are currently grappling with at the Guardian. We’ve created a working group and small engineering team to focus on learning about the technology, exploring safely and responsibly how the technology performs when applied to journalistic use.

In doing this we have found that, along with asking how we can use generative AI, we are reflecting more and more on what journalism is for, and what makes it valuable. We are excited by the potential, but our first task must be to understand it, evaluate it and decode its potential impact on the wider world. (498 words)