

Nearly 50 news websites are 'AI-generated', a study says. Would I be able to tell?

Publication Date: 2023-05-08

Author: Matthew Cantor

Section: Technology

Tags: Artificial intelligence (AI), ChatGPT, Newspapers & magazines, US press and publishing, features

Article URL: <https://www.theguardian.com/technology/2023/may/08/ai-generated-news-websites-study>



Breaking news from celebritiesdeaths.com: the president is dead. At least that's what the highly reliable website informed its readers last month, under the no-nonsense headline "Biden dead. Harris acting president, address 9am ET". The site explained that Joe Biden had "passed away peacefully in his sleep" and Kamala Harris was taking over, above a bizarre disclaimer: "I'm sorry, I cannot complete this prompt as it goes against OpenAI's use case policy on generating misleading content." Celebritiesdeaths.com is among 49 supposed news sites that NewsGuard, an organization tracking misinformation, has identified as "almost entirely written by artificial intelligence software". The sites publish up to hundreds of articles daily, according to the report, much of that material containing signs of AI-generated content, including "bland language and repetitive phrases". Some of the articles contain false information and many of the sites are packed with ads, suggesting they're intended to make money via programmatic, or algorithmically generated, advertising. The sources of the stories aren't clear: many lack bylines or use fake profile photos. In other words, NewsGuard says, experts' fears that entire news organizations could be generated by AI have already become reality. It's hard to imagine who would believe this stuff – if Biden had died, the New York Times would probably cover it – and all 49 sites contain at least one instance of AI error messaging containing phrases such as "I cannot complete this prompt" or "as an AI language model". But, as Futurism points out, a big concern here is that false information on the sites could serve as the basis for future AI content, creating a vicious cycle of fake news. What do these sites look like – and would AI articles always be as easy to spot as the report of Biden's death? I spent an afternoon in the brave new world of digital nonsense to find out. ••• The first stop was Get Into Knowledge, which offers a huge amount of knowledge to get into, all of it regurgitated on to the homepage seemingly at random. (We won't link to the sites here to avoid boosting them further.) The headlines seemed like the work of translation software. One category was "amazing reasons behind": for instance, a lengthy article on "Why do dogs eat grass? – amazing reasons behind" and "Why is yawning contagious? – 10 Amazing Science Facts behind". A piece on whether oceans freeze was based on "Massive science", and the site dares to ask questions such as "why is the Sky Blue but the Space black?" and the even more poetic "Does the gravity of Mars the same as Earth's?", something I've often wondered. I started to wonder if the language was too odd to be the work of ChatGPT, which tends to be readable, if boring. That was the case with the articles themselves. They're ordered like presentations, with an outline at the top and paragraphs arranged by number. But there are glimpses of true humanity: for instance, the piece on grass-eating dogs refers to them as our "furry friends" six times. These pieces certainly read like the work of AI, and a person who identified himself to NewsGuard as the site's founder said the site used "automation at some points where they are extremely needed". (The site did not immediately reply to

emails from the Guardian.) Once I'd gotten into enough knowledge, I visited celebritiesdeaths.com, which earnestly describes itself as "news on famous figures who have died" – a refreshing change from outlets like Us Weekly that insist on covering figures who are still alive. Other than the Biden snafu, the deaths that I factchecked had actually occurred, though they appear to have stopped in March: links to deaths in April and May didn't work. Fortunately, the shortage of deaths in those months was balanced by individuals' repeated deaths in March: the last surviving Czech second world war RAF pilot, for instance, apparently died on both the 25th and the 26th. I also learned that a "dumpling empire founder" died on 26 March, which was impressive information given that the article claimed to have been posted on 26 February. Celebritiesdeaths.com did not deem it necessary to provide the name of the founder of the "colossal global dumpling franchise", even though the 96-year-old's "demise" was widely mourned. (The piece must have referred to Yang Bing-yi, who founded a celebrated Taiwanese chain.) A Guardian email to the address listed on the site was immediately returned with an error message. Once I'd had enough of dead celebrities, I headed to ScoopEarth.com, which provides juicy insider information on stars who are still breathing, as well as, for some reason, tech tips. The first article was about the musician August Alsina, who, I learned, was born on 3 September 1992 "at the age of 30". His 3 September birthday presumably explains why "every September, Alsina has a birthday party on September 3". In an email, Niraj Kumar, identified on the site as its founder, rejected claims the site used AI, calling the material "purely genuine". Many of the pieces on the site felt too oddly worded to be ChatGPT, but there was so much repeated information that it also felt like it couldn't be written by humans. I found myself wondering how we can trust anything on the internet if it's already so difficult to tell when AI is involved. Finally, I visited Famadillo.com for product reviews. This immaculately curated site is laser-focused on stress-release tablets, RVing tips, Mother's Day T-shirts and the "top" sites in Santa Fe. The reviews themselves are sensible enough, but navigating the site is virtually impossible. Perhaps it's perfectly designed for a true dilettante – the kind of person who'd read a review of Play-Doh's Super Stretchy Green Slime immediately after a piece tackling the thorny question "Are baby potatoes regular potatoes?" In an email to the Guardian, Famadillo rejected claims it used AI to generate content highlighted in the NewsGuard report. "Famadillo runs reported interviews and reviews and uses press releases for our contest pages. None of this content is generated by AI," the email read. "That being said, we have experimented with AI in terms of refreshing old content and editing reporter-written content with the supervision of our editors." The controversy points to the growing difficulty of discerning the humans from the bots. By the end of the day, I was even more confused about what was real and what wasn't than I am after waking from a dream or watching 15 minutes of Fox News. Who, exactly, is running these sites is unclear: many don't contain contact information, and of those that NewsGuard managed to contact, most failed to reply while those that did were vague about their operations. Meanwhile, their impact appears to vary widely – some post to Facebook pages with tens of thousands of followers while others have none. If this is what AI generates now, imagine what it will look like when sites like this become AI's source material. We can only hope that the bots remain compulsively honest about their identities – or that Joe Biden finds a way to prevent an AI wild west. Assuming he's still alive.