

GUEST ESSAY

A.I.-Generated Garbage Is Polluting Our Culture

March 29, 2024

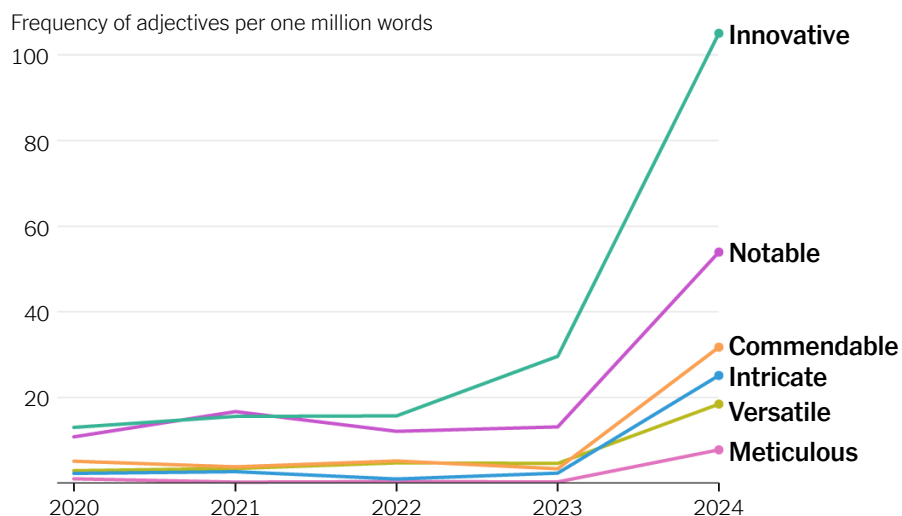
By Erik Hoel

Mr. Hoel is a neuroscientist and novelist and the author of The Intrinsic Perspective newsletter.

Increasingly, mounds of synthetic A.I.-generated outputs drift across our feeds and our searches. The stakes go far beyond what's on our screens. The entire culture is becoming affected by A.I.'s runoff, an insidious creep into our most important institutions.

Consider science. Right after the blockbuster release of GPT-4, the latest artificial intelligence model from OpenAI and one of the most advanced in existence, the language of scientific research began to mutate. Especially within the field of A.I. itself.

Adjectives associated with A.I.-generated text have increased in peer reviews of scientific papers about A.I.



Note: Peer reviews are for the International Conference on Learning Representations (ICLR), one of the largest A.I. conferences. • Source: "Monitoring AI-Modified Content at Scale: A Case Study on the Impact of ChatGPT on AI Conference Peer Reviews" • By Taylor Maggiamo

A study published this month examined scientists' peer reviews — researchers' official pronouncements on others' work that form the bedrock of scientific progress — across a number of high-profile and prestigious scientific conferences studying A.I. At one such conference, those

peer reviews used the word “meticulous” more than 34 times as often as reviews did the previous year. Use of “commendable” was around 10 times as frequent, and “intricate,” 11 times. Other major conferences showed similar patterns.

Such phrasings are, of course, some of the favorite buzzwords of modern large language models like ChatGPT. In other words, significant numbers of researchers at A.I. conferences were caught handing their peer review of others’ work over to A.I. — or, at minimum, writing them with lots of A.I. assistance. And the closer to the deadline the submitted reviews were received, the more A.I. usage was found in them.

Sign up for the Opinion Today newsletter Get expert analysis of the news and a guide to the big ideas shaping the world every weekday morning. [Get it sent to your inbox.](#)

If this makes you uncomfortable — especially given A.I.’s current unreliability — or if you think that maybe it shouldn’t be A.I.s reviewing science but the scientists themselves, those feelings highlight the paradox at the core of this technology: It’s unclear what the ethical line is between scam and regular usage. Some A.I.-generated scams are easy to identify, like the medical journal paper featuring a cartoon rat sporting enormous genitalia. Many others are more insidious, like the mislabeled and hallucinated regulatory pathway described in that same paper — a paper that was peer reviewed as well (perhaps, one might speculate, by another A.I.?).

What about when A.I. is used in one of its intended ways — to assist with writing? Recently, there was an uproar when it became obvious that simple searches of scientific databases returned phrases like “As an A.I. language model” in places where authors relying on A.I. had forgotten to cover their tracks. If the same authors had simply deleted those accidental watermarks, would their use of A.I. to write their papers have been fine?

What’s going on in science is a microcosm of a much bigger problem. Post on social media? Any viral post on X now almost certainly includes A.I.-generated replies, from summaries of the original post to reactions written in ChatGPT’s bland Wikipedia-voice, all to farm for follows. Instagram is filling up with A.I.-generated models, Spotify with A.I.-generated songs. Publish a book? Soon after, on Amazon there will often appear A.I.-generated “workbooks” for sale that supposedly accompany your book (which are incorrect in their content; I know because this happened to me). Top Google search results are now often A.I.-generated images or articles. Major media outlets like Sports Illustrated have been creating A.I.-generated articles attributed to equally fake author profiles. Marketers who sell search engine optimization methods openly brag about using A.I. to create thousands of spammed articles to steal traffic from competitors.

Then there is the growing use of generative A.I. to scale the creation of cheap synthetic videos for children on YouTube. Some example outputs are Lovecraftian horrors, like music videos about parrots in which the birds have eyes within eyes, beaks within beaks, morphing unfathomably while singing in an artificial voice, “The parrot in the tree says hello, hello!” The narratives make no sense, characters appear and disappear randomly, and basic facts like the names of shapes are wrong. After I identified a number of such suspicious channels on my newsletter, *The Intrinsic Perspective*, *Wired* found evidence of generative A.I. use in the production pipelines of some accounts with hundreds of thousands or even millions of subscribers.

As a neuroscientist, this worries me. Isn't it possible that human culture contains within it cognitive micronutrients — things like cohesive sentences, narrations and character continuity — that developing brains need? Einstein supposedly said: “If you want your children to be intelligent, read them fairy tales. If you want them to be very intelligent, read them more fairy tales.” But what happens when a toddler is consuming mostly A.I.-generated dream-slop? We find ourselves in the midst of a vast developmental experiment.

There's so much synthetic garbage on the internet now that A.I. companies and researchers are themselves worried, not about the health of the culture, but about what's going to happen with their models. As A.I. capabilities ramped up in 2022, I wrote on the risk of culture's becoming so inundated with A.I. creations that when future A.I.s are trained, the previous A.I. output will leak into the training set, leading to a future of copies of copies of copies, as content became ever more stereotyped and predictable. In 2023 researchers introduced a technical term for how this risk affected A.I. training: model collapse. In a way, we and these companies are in the same boat, paddling through the same sludge streaming into our cultural ocean.

With that unpleasant analogy in mind, it's worth looking to what is arguably the clearest historical analogy for our current situation: the environmental movement and climate change. For just as companies and individuals were driven to pollute by the inexorable economics of it, so, too, is A.I.'s cultural pollution driven by a rational decision to fill the internet's voracious appetite for content as cheaply as possible. While environmental problems are nowhere near solved, there has been undeniable progress that has kept our cities mostly free of smog and our lakes mostly free of sewage. How?

Before any specific policy solution was the acknowledgment that environmental pollution was a problem in need of outside legislation. Influential to this view was a perspective developed in 1968 by Garrett Hardin, a biologist and ecologist. Dr. Hardin emphasized that the problem of pollution was driven by people acting in their own interest, and that therefore “we are locked into a system of ‘fouling our own nest,’ so long as we behave only as independent, rational, free-enterprisers.”

He summed up the problem as a “tragedy of the commons.” This framing was instrumental for the environmental movement, which would come to rely on government regulation to do what companies alone could or would not.

Once again we find ourselves enacting a tragedy of the commons: short-term economic self-interest encourages using cheap A.I. content to maximize clicks and views, which in turn pollutes our culture and even weakens our grasp on reality. And so far, major A.I. companies are refusing to pursue advanced ways to identify A.I.’s handiwork — which they could do by adding subtle statistical patterns hidden in word use or in the pixels of images.

A common justification for inaction is that human editors can always fiddle around with whatever patterns are used if they know enough. Yet many of the issues we’re experiencing are not caused by motivated and technically skilled malicious actors; they’re caused mostly by regular users’ not adhering to a line of ethical use so fine as to be nigh nonexistent. Most would be uninterested in advanced countermeasures to statistical patterns enforced into outputs that should, ideally, mark them as A.I.-generated.

That’s why the independent researchers were able to detect A.I. outputs in the peer review system with surprisingly high accuracy: They actually tried. Similarly, right now teachers across the nation have created home-brewed output-side detection methods, like adding hidden requests for patterns of word use to essay prompts that appear only when copied and pasted.

In particular, A.I. companies appear opposed to any patterns baked into their output that can improve A.I.-detection efforts to reasonable levels, perhaps because they fear that enforcing such patterns might interfere with the model’s performance by constraining its outputs too much — although there is no current evidence this is a risk. Despite public pledges to develop more advanced watermarking, it’s increasingly clear that the companies are dragging their feet because it goes against the A.I. industry’s bottom line to have detectable products.

To deal with this corporate refusal to act we need the equivalent of a Clean Air Act: a Clean Internet Act. Perhaps the simplest solution would be to legislatively force advanced watermarking intrinsic to generated outputs, like patterns not easily removable. Just as the 20th century required extensive interventions to protect the shared environment, the 21st century is going to require extensive interventions to protect a different, but equally critical, common resource, one we haven’t noticed up until now since it was never under threat: our shared human culture.

Erik Hoel is a neuroscientist, a novelist and the author of *The Intrinsic Perspective* newsletter.

The Times is committed to publishing a diversity of letters to the editor. We’d like to hear what you think about this or any of our articles. Here are some tips. And here’s our email: letters@nytimes.com.

Follow the *New York Times* Opinion section on Facebook, Instagram, TikTok, WhatsApp, X and Threads.

