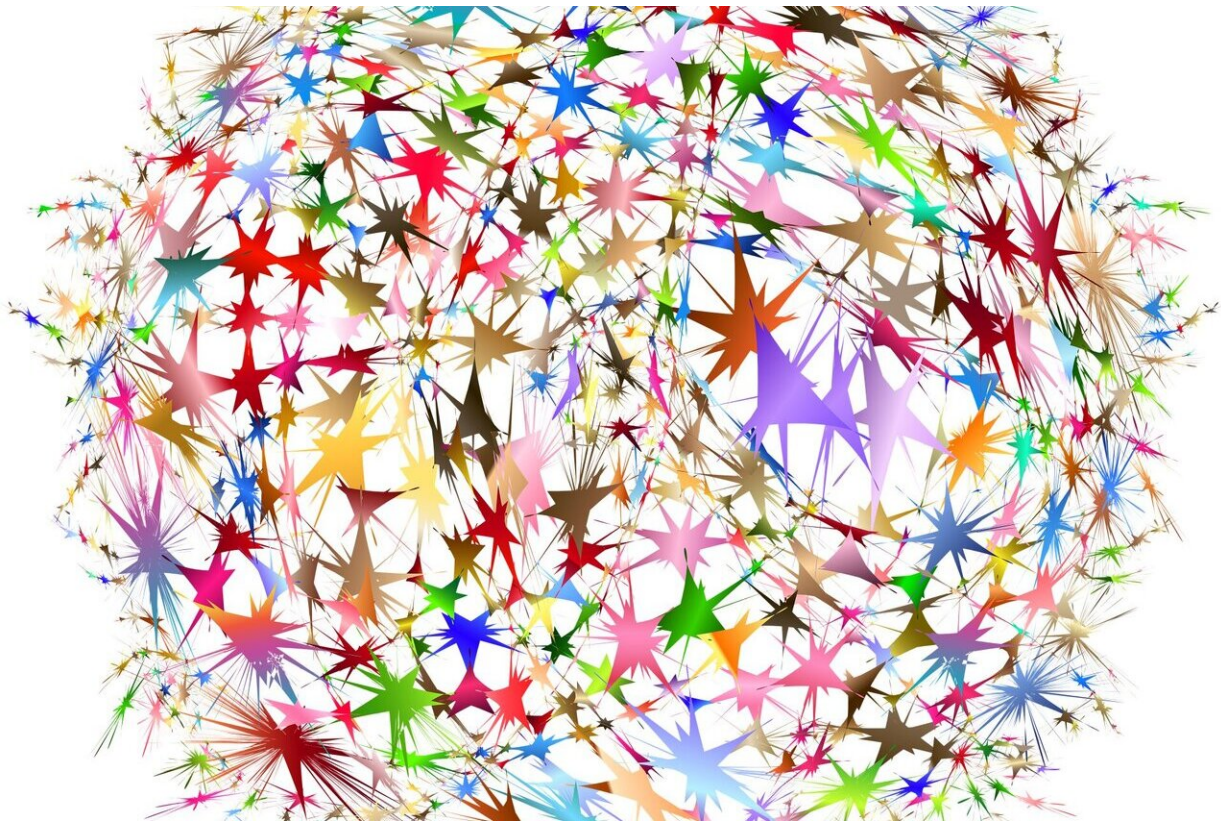# Fake news via OpenAI: Eloquently incoherent?

November 9 2019, by Nancy Cohen

Credit: CC0 Public Domain

OpenAI's text generator, machine learning-powered—so powerful that it was thought too dangerous to release to the public, has, guess what, been released.

OpenAI published a [blog](#) post announcing its decision to release the algorithm in full as it has "seen no strong evidence of misuse so far."

Well, that was a turnaround.

It was only back in February when OpenAI [talked about](#) a language model called GPT-2 that generates paragraphs of [text](#).

[Engadget](#): "The AI, GPT-2, was originally designed to answer questions, summarize stories and translate texts. But researchers came to fear that it could be used to pump out large volumes of misinformation."

Tom McKay in *Gizmodo* [noted](#) how the text generator was "trained on some 40 gigabytes of data yanked from eight million websites" and good at generating text from a given prompt.

The Verge's James Vincent also talked about it. "The system was trained on eight million text documents scraped from the web and responds to text snippets supplied by users. Feed it a fake headline, for [example](#), and it will write a [news story](#); give it the first line of a poem and it'll supply a whole verse."

So why has their algorithm been released in full? Doesn't the public have enough to worry about [fake news](#), fake criticism, political propaganda?

Besides, the February [blog](#) post took what appeared to be such an impressive high road in showing the world ethical considerations were top of mind.

"Our model, called GPT-2 (a successor to GPT), was trained simply to predict the next word in 40GB of Internet text. Due to our concerns about malicious applications of the technology, we are not releasing the trained model. As an experiment in responsible disclosure, we are

instead releasing a much smaller model for researchers to experiment with, as well as a technical paper."

Katyanna Quach in *The Register* [provided](#) a timeline of the OpenAI tech appearances in stages. "The San Francisco-based research lab tentatively tested the waters by releasing larger and larger models, starting from just a few hundred million parameters.

"The smallest version contained 117 million parameters, the second had 345 million parameters, the third consisted of 774 million parameters, and the largest one, released on Tuesday, has the full 1.5 billion parameters. The more parameters, the more powerful and capable the model, generally speaking."

So, what convinced them to take the exit ramp off the high road? What reason did they have to think that now was a safer time than ever before? What happened to the position that it was too dangerous to share?

"We've seen no strong evidence of misuse so far," they stated in their Nov. 5 post. "While we've seen some discussion around GPT-2's potential to argument high-volume/low-yield operations like spam and phishing, we haven't seen evidence of writing code, documentation, or instances of misuse."

Meanwhile, James Vincent in *The Verge* ventured to air the possibility that this "dangerous" notion was just that, a notion that, as far as their fake-news model, may not have been called for in the first place.

There was concern that easy access to such an AI tool could satisfy the plans of malicious actors, and the responsible way to handle this was to release the model in a limited fashion. A good approach? Or was it thwarting further research? "

You can now try it out on a [site](#), which runs the full-sized GPT-2 [model](#), called 1558M. The web interface was built by Adam King using the full GPT-2 release from OpenAI, said *The Register*.

This writer tried to explore what it can do with a starter sentence "Babies can wash the dishes." What followed was a zany mishmash of logic and topic leaps. Nonetheless the sentences carried respectable grammar, voice and style. Some snippets:

"And so, this idea that children are incapable of using a washing machine is ridiculous.

"As far as the idea that your kid is a sink faucet, that's an idea that I personally don't think I can stand. I've been known to use a sink faucet when I'm cooking. I can wash dishes in a sink.

"When your child is sitting on your lap and you're holding the washcloth and using the water to wash the dishes, you're probably not doing anything."

James Vincent hit the mark with a three word assessment: "eerily cogent writing."

Quach also made a point worth repeating. "Occasionally, it spits out sentences that are surprisingly good, but as it keeps churning out text, it becomes incoherent."

Quach had a refreshing reason why OpenAI has contributed to humanity. "We have one final thought on this fascinating AI research: it's at least set a bar for human writers. If you want to write news or feature articles, blog posts, marketing emails, and the like, know that you now have to be better than GPT-2's semi-coherent output. Otherwise, people might as well just read a bot's output than your own."

More information: openai.com/blog/gpt-2-1-5b-release/

Citation: Fake news via OpenAI: Eloquently incoherent? (2019, November 9) retrieved 10 April 2024 from https://techxplore.com/news/2019-11-fake-news-openai-eloquently-incoherent.html