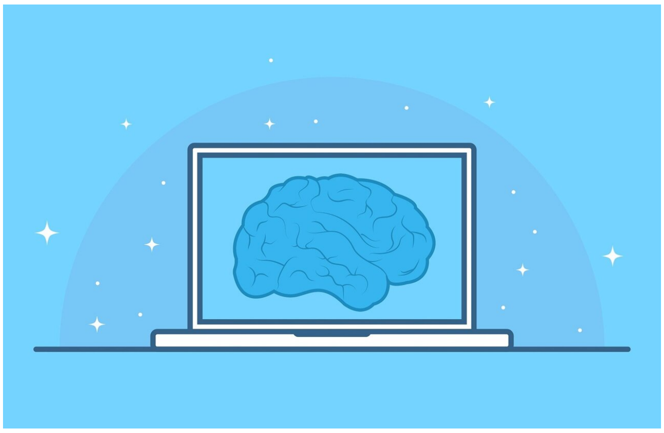
Fake news model in staged release but two researchers fire up replication

31 August 2019, by Nancy Cohen



Credit: CC0 Public Domain

Not the most comforting news in the world of tech: The artificial intelligence lab (OpenAI) cofounded by Elon Musk said its software could too easily be adapted to crank out fake news. "Two grads re-created it anyway." That was *Wired*'s coverage on August 26 of a story about two recent master's graduates in computer science having released what they said was "a re-creation of OpenAI's withheld software" for anyone to download and use.

Withheld? Why? It had been withheld over concerns about the societal impact.

In February, OpenAI announced their model, GPT-2, and said it was trained to predict the next word in 40GB of Internet text.

They spelled out their release strategy: "Due to concerns about large language models being used to generate deceptive, biased, or abusive language at scale, we are only releasing a much smaller version of GPT-2 along with sampling code. We are not releasing the dataset, training code, or GPT-2 model weights." In May, said *MIT Technology Review*, "a few months after GPT-2's initial debut, OpenAI revised its stance on withholding the full code to what it calls a "staged release."

Charanjeet Singh in *Fossbytes* said that the software analyzed language patterns and could be used up for tasks like chatbots and coming up with unprecedented answers but "the most alarming concern among experts has been the creation of synthetic text."

Well, the two grads in the news released a re-creation of the OpenAI software onto the Internet but the two researchers, Aaron Gokaslan ad Vanya Cohen, never wanted to drain oceans or make the sky fall.

Tom Simonite, who wrote the much quoted article in *Wired*, said the two researchers, ages 23 and 24, were not out to cause havoc but said their release was intended to show that you don't have to be an elite lab rich in dollars and Ph.D.s to create this kind of software: They used an estimated $50,000 worth of free cloud computing from Google.

Sissi Cao, *Observer*: Similar to OpenAI's process, Gokaslan and Cohen trained their language software using webpages of text "written by humans (by harvesting links shared on Reddit) and cloud computing from Google.

What is more, the researchers' actions being potentially dangerous could be debated.

Simonite made this point: "Machine learning software picks up the statistical patterns of language, not a true understanding of the world. Text from both the original and wannabe software often makes nonsensical leaps. Neither can be directed to include particular facts or points of view."

Sample output was provided by Gokaslan and

Cohen in *Medium* and, for sure, it is a head-scratcher as one attempts to find any logic flow from one sentence to another.

That article was titled "OpenGPT-2: We Replicated GPT-2 Because You Can Too." They said they believed releasing their model was a reasonable first step towards countering the potential future abuse of these kinds of models. He said they modified their codebase to match the language modeling training objective of GPT-2. "Since their model was trained on a similarly large corpus, much of the code and hyper-parameters proved readily reusable."

Since Open-AI had not released their largest model at this time [the date of his posting was August 22], he said the two researchers sought to replicate their 1.5B model to allow others to build on their pretrained model and further improve it.

Fast forward to August 29. Where does all this leave OpenAI's GPT-2? Karen Hao in *MIT Technology Review* said that he policy team has published a paper, submitted on 24 Aug, which is now up on arXiv, and "Alongside it, the lab has released a version of the model, known as GPT-2, that's half the size of the full one, which has still not been released."

Hao's article was particularly useful in understanding this fake-text drama as she reported on how the staged-release approach was being received outside of OpenAI.

A deep learning engineer at Nvidia said he did not think a staged release was particularly useful in this case because the work was easily replicable, "But it might be useful in the way that it sets a precedent for future projects. People will see staged release as an alternative option."

She also quoted Oren Etzioni, the CEO of the Allen Institute for Artificial Intelligence. "I applaud their intent to design a thoughtful, gradual release process for AI technology but question whether all the fanfare was warranted."

More information: openai.com/blog/better-language-models/

APA citation: Fake news model in staged release but two researchers fire up replication (2019, August 31) retrieved 25 January 2021 from https://techxplore.com/news/2019-08-fake-news-staged-replication.html