OpenAI's new language generator GPT-3 is shockingly good—and completely mindless

technologyreview.com/2020/07/20/1005454/openai-machine-learning-language-generator-gpt-3-nlp/

Will Douglas Heaven

"Playing with GPT-3 feels like seeing the future," Arram Sabeti, a San Francisco–based developer and artist, tweeted last week. That pretty much sums up the response on social media in the last few days to OpenAI's latest language-generating AI.

OpenAI first described GPT-3 in a research paper published in May. But last week it began drip-feeding the software to selected people who requested access to a private beta. For now, OpenAI wants outside developers to help it explore what GPT-3 can do, but it plans to turn the tool into a commercial product later this year, offering businesses a paid-for subscription to the AI via the cloud.

> Working towards the future where all of the internet is a simulacrum of previous versions of the internet https://t.co/2dIjrlKNVG
>
> — Julian Togelius (@togelius) July 19, 2020

GPT-3 is the most powerful language model ever. Its predecessor, GPT-2, released last year, was already able to spit out convincing streams of text in a range of different styles when prompted with an opening sentence. But GPT-3 is a big leap forward. The model has 175 billion parameters (the values that a neural network tries to optimize during training), compared with GPT-2's already vast 1.5 billion. And with language models, size really does matter.

Stay updated on MIT Technology Review initiatives and events?

Sabeti linked to a blog post where he showed off short stories, songs, press releases, technical manuals, and more that he had used the AI to generate. GPT-3 can also produce pastiches of particular writers. Mario Klingemann, an artist who works with machine learning, shared a short story called "The importance of being on Twitter," written in the style of Jerome K. Jerome, which starts: "It is a curious fact that the last remaining form of social life in which the people of London are still interested is Twitter. I was struck with this curious fact when I went on one of my periodical holidays to the sea-side, and found the whole place twittering like a starling-cage." Klingemann says all he gave the AI was the title, the author's name and the initial "It." There is even a reasonably informative article about GPT-3 written entirely by GPT-3.

> Another attempt at a longer piece. An imaginary Jerome K. Jerome writes about Twitter. All I seeded was the title, the author's name and the first "It", the rest is done by #gpt3
>
> Here is the full-length version as a PDF:https://t.co/d2gpmlZ1T5 pic.twitter.com/1N0lNoC1eZ
>
> — Mario Klingemann (@quasimondo) July 18, 2020

Others have found that GPT-3 can generate any kind of text, including guitar tabs or computer code. For example, by tweaking GPT-3 so that it produced HTML rather than natural language, web developer Sharif Shameem showed that he could make it create web-page layouts by giving it prompts like “a button that looks like a watermelon” or “large text in red that says WELCOME TO MY NEWSLETTER and a blue button that says Subscribe.” Even legendary coder John Carmack, who pioneered 3D computer graphics in early video games like Doom and is now consulting CTO at Oculus VR, was unnerved: “The recent, almost accidental, discovery that GPT-3 can sort of write code does generate a slight shiver.”

> This is mind blowing.
>
> With GPT-3, I built a layout generator where you just describe any layout you want, and it generates the JSX code for you.
>
> W H A T pic.twitter.com/w8JkrZO4lk
>
> — Sharif Shameem (@sharifshameem) July 13, 2020

Yet despite its new tricks, GPT-3 is still prone to spewing hateful sexist and racist language. Fine-tuning the model helped limit this kind of output in GPT-2.

> I mean, maybe Iâ€™m just jaded but Iâ€™m going to wait a bit and see what sort of egregious bias comes out of GPT-3. Oh, it writes poetry? Nice. Oh, it also spews out harmful sexism and racism? I am rehearsing my shocked face. #gpt3
>
> — Kate Devlin (@drkatedevlin) July 18, 2020

> No one should be surprised by this. How do we keep this from happening accidentally? Donâ€™t have all the answers yet, but fine-tuning on strong and generalizable normative priors helped with GPT-2 https://t.co/V12NM8ZtAH https://t.co/1bn6G6eWjM
>
> — Mark Loki Variant Riedl (@mark_riedl) July 18, 2020

It’s also no surprise that many have been quick to start talking about intelligence. But GPT-3’s human-like output and striking versatility are the results of excellent engineering, not genuine smarts. For one thing, the AI still makes ridiculous howlers that reveal a total lack of

common sense. But even its successes have a lack of depth to them, reading more like cut-and-paste jobs than original compositions.

> This post is one of the best GPT-3 evaluations I've seen. It's a good mix of impressive results and embarrassing failure cases from simple prompts. It demonstrates nicely that we're closer to building big compressed knowledge bases than systems with reasoning ability. https://t.co/a5Nq006dMD
>
> — Denny Britz (@dennybritz) July 17, 2020

> This supports my suspicion that GPT-3 uses a lot of its parameters to memorize bits of text from the internet that donâ€™t generalize easily https://t.co/I7uS4iu2sn
>
> — Mark Loki Variant Riedl (@mark_riedl) July 19, 2020

Exactly what's going on inside GPT-3 isn't clear. But what it seems to be good at is synthesizing text it has found elsewhere on the internet, making it a kind of vast, eclectic scrapbook created from millions and millions of snippets of text that it then glues together in weird and wonderful ways on demand.

> GPT-3 often performs like a clever student who hasn't done their reading trying to bullshit their way through an exam. Some well-known facts, some half-truths, and some straight lies, strung together in what first looks like a smooth narrative.
>
> — Julian Togelius (@togelius) July 17, 2020

That's not to downplay OpenAI's achievement. And a tool like this has many new uses, both good (from powering better chatbots to helping people code) and bad (from powering better misinformation bots to helping kids cheat on their homework).

But when a new AI milestone comes along it too often gets buried in hype. Even Sam Altman, who co-founded OpenAI with Elon Musk, tried to tone things down: "The GPT-3 hype is way too much. It's impressive (thanks for the nice compliments!) but it still has serious weaknesses and sometimes makes very silly mistakes. AI is going to change the world, but GPT-3 is just a very early glimpse. We have a lot still to figure out."

We have a low bar when it comes to spotting intelligence. If something looks smart, it's easy to kid ourselves that it is. The greatest trick AI ever pulled was convincing the world it exists. GPT-3 is a huge leap forward—but it is still a tool made by humans, with all the flaws and limitations that implies.

> Seeing so much on Twitter about GPT-3. Remember...
>
> The Turing Test is not for AI to pass, but for humans to fail.
>
> — Mark Loki Variant Riedl (@mark_riedl) July 18, 2020