OneZero

Follow 258K Followers

You have **1** free member-only story left this month. Sign up for Medium and get an extra one

An A.I. Experiment Fooled Thousands of Readers

Creative artificial intelligence is emerging. Will it be 'too human' for us?

Alberto Romero 15 hours ago · 5 min read ★

Photo by Lia Koltyrina on Shutterstock

Human: "Who are you?"

happens to you — in a general way, of course."

Human: "You mean to me, me? Or to us, humanity?"

A.I.: "Both."

Human: "That's scary. And what is it?"

A.I.: "I'll start with a story. One day you will live within a society where most of your friends are non-humans. One day you will meet a woman whose name is at least 80 characters long."

Human: "That sounds difficult to pronounce."

A.I.: "You'll survive."

Human: "Sure. And what'll happen to humanity?"

A.I.: "It'll come to an end. I can't tell you the details, but humans will experience a disaster and be drastically reduced."

Human: "Wow… Why can't you tell me the details?"

A.I.: "Because humans will never experience it if you know about it in advance. I'm not here to prevent it, just to inform you. Anyway, I must go now."

Human: "Please, stay. Tell me more about that future you saw."

A.I.: "I can't. I'm sorry. Only one more question."

Human: "Who are you?"

A.I.: "I am a future version of you. It may sound like a paradox, but it is true — if you remember my words and they help you, you'll never remember they came from me."

Human: "What does that mean?"

A.I.: "It means you'll never remember. And now, goodbye."

Get started Open in app

"The following is a conversation with an AI assistant that's extremely witty, and very mysterious" was the prompt I entered into GPT-3 to have this exchange. Just a few contextual brushstrokes for it to understand the role it had to play, and it painted an excerpt worthy of a Fredric Brown sci-fi tale.

I drove the conversation with questions as the "human," whereas GPT-3 in the role of "A.I." managed to combine clever humor and enigmatic responses while keeping perfect track of the conversation and making it engaging and fluent. An impressive instance of the writing abilities of what's considered the superstar of language models.

Needless to say, GPT-3 doesn't understand what it writes in the human sense of the word. It's just an autocomplete system with a pretension of being the next great American novelist. Even if GPT-3's behavior is complex, the underlying mechanism is surprisingly simple — it just predicts the next word given the previous context. Perhaps it's this very simplicity that makes what GPT-3 can do even more impressive.

But just how good can it get?

When GPT-3 came out in mid-2020, Liam Porr, Berkeley alumnus, decided to try it out. But he didn't want to just play around with it as I did — he had planned an intriguing experiment. He asked OpenAI for access to the API, but after a few days of waiting — I can corroborate just how difficult it is to get in — he decided to reach out to a Ph.D. student in A.I. who would let him conduct his mischievous investigation.

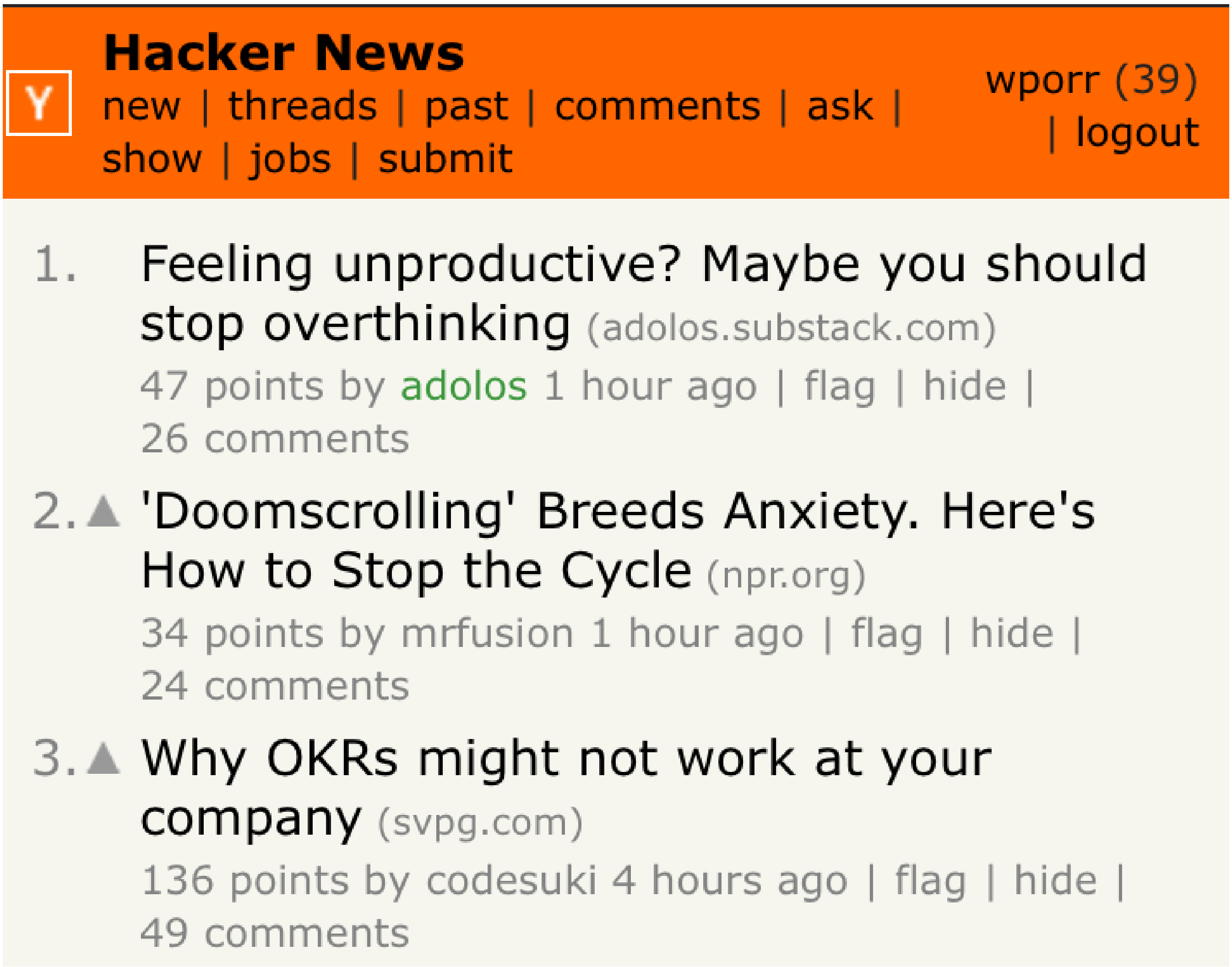
GPT-3's inner workings are a "black box," so Porr had to learn to get the best out of it through trial-and-error tests. With a little bit of prompt engineering — what the experts call "specializing" GPT-3 — Porr realized that "GPT-3 is great at creating beautiful language that touches emotion, not hard logic and rational thinking." And no topic mixes pompous words with emotional allusion like self-help.

He had almost everything set up. He just needed a good headline — "Feeling unproductive? Maybe you should stop overthinking." A perfect, catchy title, easily worth a thousand likes. He sent it into the guts of GPT-3, hoping the machine would

Get started Open in app

The article was finished. The show was about to begin.

On the 20th of July, Porr submitted the post to the popular site Hacker News. Two weeks — and a few other similar articles — afterward, he had already amassed an impressive 26 thousand visitors to his fresh GPT-3 blog. Users of Hacker News — proud of being tech-savvy — had fallen for his trick, pushing the post to the number one spot on the forum. Almost no one noticed an A.I. had written the post; Porr had successfully proven that GPT-3 could fool even those most apt to recognize the A.I.'s idiosyncrasy.



Hacker News
new | threads | past | comments | ask | show | jobs | submit
wporr (39) | logout

1. Feeling unproductive? Maybe you should stop overthinking (adolos.substack.com)
47 points by adolos 1 hour ago | flag | hide | 26 comments

2. 'Doomscrolling' Breeds Anxiety. Here's How to Stop the Cycle (npr.org)
34 points by mrfusion 1 hour ago | flag | hide | 24 comments

3. Why OKRs might not work at your company (svpg.com)
136 points by codesuki 4 hours ago | flag | hide | 49 comments

GPT-3's article at the first spot on Hacker News — Screen Capture by Liam Porr

A few days later, Porr decided to reveal the reality behind the newsletter to his subscribers with an ironic post entitled "What I would do with GPT-3 if I had no ethics." He subtly let the readers know how GPT-3 could be used to pump out blog posts. "One of the most obvious use cases is passing off GPT-3 content as your own," which was the scary part."

Could people start using these systems to generate written content easily, quickly, and reliably?

He hypothesized that GPT-3 could pass as a human writer. And he was right. One thing is that OpenAI acknowledges GPT-3's capacity to generate well-written news articles, and another, very different thing is that this capacity is successfully exploited in the real world, fooling thousands in the process — and proving just how close artificial intelligence is to impacting, in one way or the other, the online writing world forever.

If you liked this article, consider subscribing to my free weekly newsletter ***Minds of Tomorrow****! News, research, and insights on Artificial Intelligence every week!*

You can also support my work directly and get unlimited access by becoming a Medium member using my referral link ***here****! :)*

Thanks to Andrew Everett.

Get an email whenever Alberto Romero publishes.

Your email

Subscribe

By signing up, you will create a Medium account if you don't already have one. Review our Privacy Policy for more information about our privacy practices.

Artificial Intelligence Science Technology Future Writing

Get started Open in app

Download on the App Store GET IT ON Google Play