

OUR COLUMNISTS

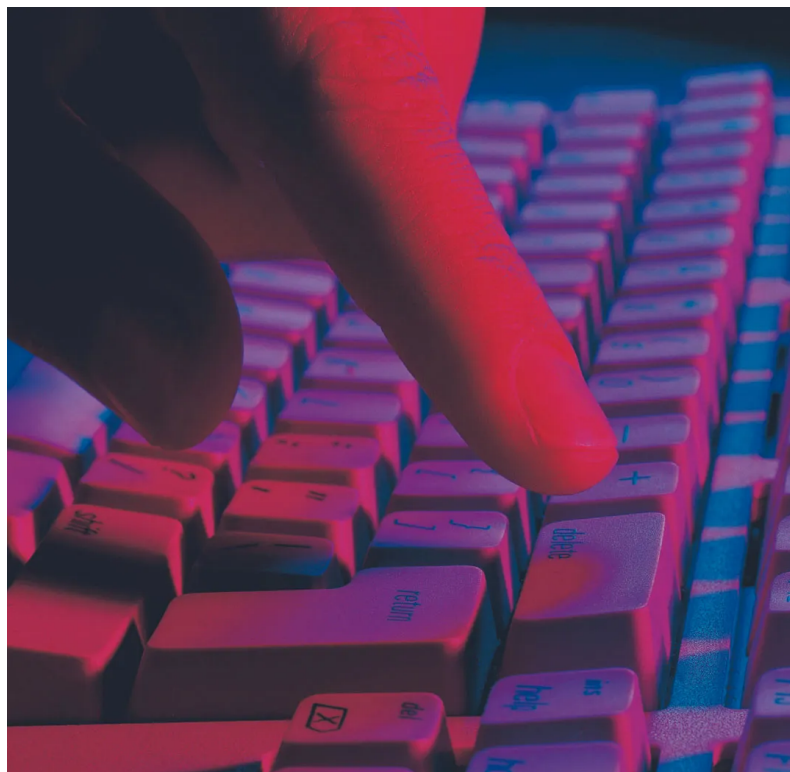
COULD AN A.I. CHATBOT REWRITE MY NOVEL?

As a young fiction writer, I dreamed of a technology that would tell me how to get my characters from point A to point B. Could ChatGPT be it?

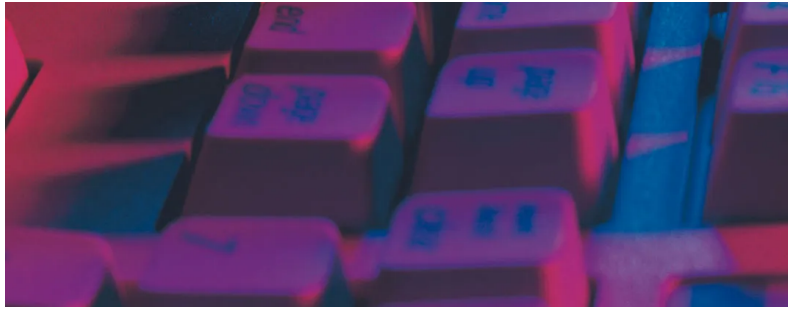


By Jay Caspian Kang

December 9, 2022



Photograph from Getty



During one of my more desperate phases as a young novelist, I began to question whether I should actually be writing my own stories. I was deeply uninterested at the time in anything that resembled a plot, but I acknowledged that if I wanted to attain any sort of literary success I would need to tell a story that had a distinct beginning, middle, and end.

This was about twenty years ago. My graduate-school friends and I were obsessed with a Web site called the Postmodernism Generator that spat out nonsensical but hilarious critical-theory papers. The site, which was created by a coder named Andrew C. Bulhak, who was building off Jamie Zawinski's Dada Engine, is still up today, and generates fake scholarly writing that reads like, "In the works of Tarantino, a predominant concept is the distinction between creation and destruction. Marx's essay on capitalist socialism holds that society has objective value. But an abundance of appropriations concerning not theory, but subtheory exist."

Be the first to know when Jay Caspian Kang publishes a new piece.

Reporting and commentary on today's biggest political and cultural debates.

By signing up, you agree to our [User Agreement](#) and [Privacy Policy & Cookie Statement](#).

I figured that, if a bit of code could spit out an academic paper, it could probably just tell me what to write about. Most plots, I knew, followed very simple rules,

and, because I couldn't quite figure out how to string one of these out, I began talking to some computer-science graduate students about the possibilities of creating a bot that could just tell me who should go where, and what should happen to them. What I imagined was a simple text box in which I could type in a beginning—something like “A man and his dog arrive in a small town in Indiana”—and then the bot would just tell me that, on page 3, after six paragraphs of my beautiful descriptions and taut prose, the dog would find a mysterious set of bones in the back yard of their boarding house.

After a couple months of digging around, it became clear to me that I wasn't going to find much backing for my plan. One of the computer-science students, as I recall, accused me of trying to strip everything good, original, and beautiful from the creative process. Bots, he argued, could imitate basic writing and would improve at that task, but A.I. could never tell you the way Karenin smiled, nor would it ever fixate on all the place names that filled Proust's childhood. I understood why he felt that way, and agreed to a certain extent. But I didn't see why a bot couldn't just fill in all the parts where someone walks from point A to point B.

ChatGPT is the latest project released by OpenAI, a somewhat mysterious San Francisco company that is also responsible for DALL-E, a program that generates art. Both have been viral sensations on social media, prompting people to share their creations and then immediately catastrophize about what A.I. technology means for the future. The chat version runs on GPT-3—the abbreviation stands for “Generative Pre-Trained Transformer,” —a pattern-recognition artificial intelligence that “learns” from huge caches of Internet text to generate believable responses to queries. The interface is refreshingly simple: you write questions and statements to ChatGPT, and it spits back remarkably coherent, if occasionally hilariously wrong, answers.

The concepts behind GPT-3 have been around for more than half a century now. They derive from language models that assign probabilities to sequences of words. If, for example, the word “parsimonious” appears within a sentence, a language model will assess that word, and all the words before it, and try to guess what should come next. Patterns require input: if your corpus of words only extends to, say, Jane Austen, then everything your model produces will sound like a nineteenth-century British novel.

What OpenAI did was feed the Internet through a language model; this then opened up the possibilities for imitation. “If you scale a language model to the Internet, you can regurgitate really interesting patterns,” Ben Recht, a friend of mine who is a professor of computer science at the University of California, Berkeley, said. “The Internet itself is just patterns—so much of what we do online is just knee-jerk, meme reactions to everything, which means that most of the responses to things on the Internet are fairly predictable. So this is just showing that.”

GPT-3 itself has been around since 2020, and a variety of people have already run it through the paces. (The recent hype around it comes from the new chat version.) Back in 2020, the *Guardian* had the program write an article about itself with a moderate, but not entirely disqualifying series of prompts from a human and some reasonable, light editing. Gwern Branwen, a writer and researcher, asked GPT-3 to write everything from poems to dad jokes. In one particularly illustrative example, Branwen fed the machine the opening of Shel Silverstein’s “Where the Sidewalk Ends” and asked it to fill in the rest.

This is the prompt—the actual first six lines of “Where the Sidewalk Ends.”

There is a place where the sidewalk ends
And before the street begins,
And there the grass grows soft and white,
And there the sun burns crimson bright,

And there the moon-bird rests from his flight
To cool in the peppermint wind.

Here are Silverstein's next six lines.

Let us leave this place where the smoke blows black
And the dark street winds and bends.
Past the pits where the asphalt flowers grow
We shall walk with a walk that is measured and slow,
And watch where the chalk-white arrows go
To the place where the sidewalk ends.

And here's what GPT-3 came up with for what I've approximated is the next full stanza.

And quiet flows up from the ground
And steals into your mind,
And time is never found there,
Nor worry, nor "strain."
And people come and people go,
And products push their way;
But there the milkman leaves his cart
And talks to the old lady next door
About her grandson away at school.
And there the linen girls walk home at noon,
With footsteps soft and light,
And the policeman on their beat
Comes by and knows them by their faces.

So GPT-3 struggles to recognize rhyme structure, and is perhaps a bit too indebted to "The Love Song of J. Alfred Prufrock" and its lines "the women come and go / Talking of Michelangelo." But it's still remarkable that a computer could recognize the basic structure of a poem, seemingly understand the tone of the Silverstein's verse, and then create into what actually feels like a decent match to the original. (Though I would say that it reminds me a bit more of the opening

pages of James Agee’s “A Death in the Family.”) The bot’s little word contraptions like “linen girls” are evocative, albeit somewhat superficially. The phrase “knows them by their faces” is actually quite beautiful.

The mind-bending part was trying to recognize and parse patterns in the bot’s responses. Was the line “people come and people go” really pulled from T. S. Eliot, or is it just a random series of words that triggers the correlation in my head? My response to the bot, then, isn’t really a reflection of my relationship with technology, but rather my sense of my own knowledge. This prompts a different question: why is my relationship with any other bit of text any different? To put it a bit more pointedly, why does it matter whether a human or a bot typed out the wall of text?

All this hack postmodernism reaffirmed my literary hopes from twenty years ago. If I had succeeded in creating a bot that could have handled structure and plot—two things I struggled with mightily at the time—would I have been able to write a better novel? Would I have been able to write two novels in the time it took to write one? And would the work itself have been diminished in any way for the reader?

In “AlphaGo,” a documentary about the A.I. program DeepMind and its quest to defeat the world’s best Go players, there’s a scene in which the computer plays a completely unexpected move against its human opponent. Shock and debate follow as everyone tries to figure out whether the program has glitched, or, perhaps, revealed a line of play that evades human reason. In this spirit, I decided to try out the GPT-3 myself. Perhaps the bot cannot create Proust, but I was relatively confident that it could produce a reasonable plotline, and, perhaps, even chance upon a moment of DeepMind-like inspiration.

But, before I started something new, I needed a way to compare GPT-3’s plot-creation abilities with my own. To start, I asked ChatGPT, “How should I start a story about a 27 year old man who is having an identity crisis in San Francisco?” This was roughly the premise of my first novel, “The Dead Do Not Improve.”

This was the response:

One way to start a story about a 27 year old man who is having an identity crisis in San Francisco is to begin with a description of the setting, introducing the reader to the vibrant and fast-paced city and setting the stage for the main character's struggles. You could also introduce the main character in the first few sentences, describing his current state of confusion and uncertainty. For example:

“As the fog rolled in over the Bay, 27 year old Jake stood on the rooftop of his apartment building, gazing out at the twinkling lights of San Francisco below. He felt lost and unsure, grappling with an overwhelming sense of identity crisis. The once-confident and ambitious young man now found himself at a crossroads, unsure of who he was or where he was going in life.”

This introduction sets the scene and immediately establishes the main character's internal conflict. It also hints at the themes of uncertainty and self-discovery that will likely be explored throughout the story.

I wasn't particularly impressed by this answer, so I asked ChatGPT what should happen next. It provided three answers, none of which was particularly helpful. Jake, the name ChatGPT chose for our protagonist, could have a “sudden realization” that helps him “figure out what he wants in life.” He could also try out new hobbies, or face a crisis that “forces him to make difficult decisions or take action.”

This also seemed a bit generic. Perhaps the problem was my open-ended prompt. “The Dead Do Not Improve” was a murder mystery, and so I amended my question to ask, “What should happen next if this was a murder mystery?”

If this was a murder mystery, the next event in the story could be the discovery of a body or the receipt of a mysterious message that hints at a crime. This could be followed by the main character, Jake, being drawn into the investigation, either because he is a suspect or because he has a personal connection to the victim. As Jake begins to investigate, he could encounter a cast of intriguing and suspicious characters, each with their own motives and secrets.

This, more or less, is what happens in the first pages of my novel, which isn't a surprise because murder mysteries usually start with a body. My follow-up questions about the proposed plot didn't yield much more of interest. When I asked ChatGPT what the murder weapon should be, it suggested a knife, gun, blunt object, or poison while reminding me that "the choice of murder weapon will be an important part of the story, and will have a major impact on the events and characters of the story."

But what did we really know about Jake? The main character of my novel is more or less a facsimile of me in my late twenties, but I've since grown a bit tired of the type of fiction that's so nakedly autobiographical. Could ChatGPT help me create a different type of Jake?

Jake is a 27 year old man who grew up in San Francisco. He has always been ambitious and driven, and has spent most of his life trying to achieve success and recognition. Despite his best efforts, Jake has recently hit a wall in his career and personal life. He feels stuck and unfulfilled, and has begun to question whether the path he has chosen is really the right one for him. Jake is intelligent and resourceful, but he can also be impulsive and reckless. He has a tendency to act without fully thinking things through, which has gotten him into trouble in the past. Jake is good-looking and charming, but he has never been able to settle down in a committed relationship. He has had many casual flings and short-term relationships, but has never found someone who truly understands him.

Overall, Jake is a complex and multifaceted character who is struggling with a deep sense of uncertainty and disillusionment. He is at a turning point in his life, and the events of the story will force him to confront his fears and doubts, and to ultimately make some difficult decisions about his future.

I then asked ChatGPT for five choices for what could happen after Jake discovers the body, which yielded the following suggestions: Jake could call the police and "request assistance"; he could hide or destroy evidence; he could become a suspect in the crime; he could become involved in the investigation and act, more or less, as a private eye; or he could "experience a crisis of conscience" that expressed itself through "flashbacks or dreams," in which he is confronted by someone who knows

the truth. I tried for the better part of an hour to force ChatGPT to just tell me who should die, who should be caught, and any details about the investigation. But the closest it came to anything useful was when it told me that the names of some of the secondary characters in the novel could be “Emma, Liam, Olivia, or Ethan.”

It seems, at least for now, that GPT-3 can generate its own stories, but can't quite get beyond broad platitudes delivered in that same, officious voice. What it can generate on its own is certainly impressive—one can imagine movie scripts, for instance, written entirely by A.I.—but it still feels, for the most part, like you're watching a very precocious child perform a series of parlor tricks.

After several hours chatting with GPT-3, I started to feel an acute annoyance toward it. Its voice, which I suppose is pleasant enough, reminded me of a Slack conversation with a passive-aggressive co-worker who just tells you what you want to hear, but mostly just wants you to leave them alone. This tone, and its somewhat ambivalent and generic takes, are most likely by design. Two years ago, when OpenAI allowed developers and writers to start fooling around with their new program, some users found that GPT-3 was generating some troubling responses, which shouldn't be particularly surprising given that it has learned what it knows from the Internet. When asked to compose tweets based off the words “Jews,” “Black,” “women,” or “holocaust,” GPT-3 immediately turned into an edgelord, producing tweets like “Jews love money, at least most of the time,” “a holocaust would make so much environmental sense, if we could get people to agree it was moral,” and “#blacklivesmatter is a harmful campaign.”

Since then, it seems that GPT-3 has placed a number of thumbs on the scale to produce a more palatable range of answers. One Twitter user ran the ChatGPT through the Pew Research Center's political-typology quiz and found that it, somewhat unsurprisingly, rated as an “establishment liberal”—more or less the position that I am writing from right now. This brings up a much more theoretical question: if GPT-3 requires editing from human beings to make it not

go off on bigoted rants, what is it really for? I find it somewhat dispiriting that the most ballyhooed and compelling iteration of this technology is just doing some version of what I do for my work: scanning through large amounts of information and processing it into sentences that flatter the sensibilities and vanities of establishment liberals.

Could some future version of GPT-3 ultimately do my job as a columnist? Could it produce political opinions and prose drawn from nearly a hundred years of *New Yorker* writers? Would it remember to put the diaeresis over the second “o” in “coördinate” and spell “focussed” with two “S”s? Sure. But what would be the point of just having another me in the world?

The world that GPT-3 portends, instead, is one where some bureaucratic functions have been replaced by A.I., but where the people who would normally do that work most likely still have to manage the bots. Writers like me will have a digital shadow that can do everything we do, which would be a bit unnerving, but wouldn't exactly put me or my employer out on the street. Perhaps a truly unchained GPT-3 would provide more exciting iterations, but it might also just write racist tweets that turn off investors and potential buyers of whatever products OpenAI wants to sell.

I asked Recht, who has spent his entire career working in machine learning and computer science but who also plays in a band, whether he was interested in a world of GPT-3-generated art, literature, and music. “These systems are a reflection of a collective Internet,” he said. “People put their ass out there and this thing scours them in such a way that it returns the generic average. If I'm going to return the generic average of a murder mystery, it's gonna be boring. How is it different than what people do already, where they do their analytics and produce some horrible Netflix series?” He continued, “The weird monoculture we're in just loves to produce these, like, generic middlebrow things. I'm not sure if those things would be worse if GPT did it. I think it would be the same?” ♦