

The Year A.I. Ate the Internet

By Sue Halpern

The New Yorker

December 8, 2023

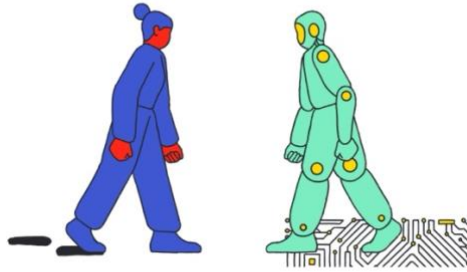


Illustration by Pavel Popov

A little more than a year ago, the world seemed to wake up to the promise and dangers of artificial intelligence when OpenAI released ChatGPT, an application that enables users to converse with a computer in a singularly human way. Within five days, the chatbot had a million users. Within two months, it was logging a hundred million monthly users—a number that has now nearly doubled. Call this the year many of us learned to communicate, create, cheat, and collaborate with robots.

Shortly after ChatGPT came out, Google released its own chatbot, Bard; Microsoft incorporated OpenAI’s model into its Bing search engine; Meta debuted LLaMA; and Anthropic came out with Claude, a “next generation AI assistant for your tasks, no matter the scale.” Suddenly, the Internet seemed nearly animate. It wasn’t that A.I. itself was new: indeed, artificial intelligence has become such a routine part of our lives that we hardly recognize it when a Netflix algorithm recommends a film, a credit-card company automatically detects fraudulent activity, or Amazon’s Alexa delivers a summary of the morning’s news.

But, while those A.I.s work in the background, often in a scripted and brittle way, chatbots are responsive and improvisational. They are also unpredictable. When we ask for their assistance, prompting them with queries about things we don’t know, or asking them for creative help, they often generate things that did not exist before, seemingly out of thin air. Poems, literature reviews, essays, research papers, and three-act plays are delivered in plain, unmistakably human language. It’s as if the god in the machine had been made in our image. Ask ChatGPT to write a song about self-driving cars in the style of Johnny Cash and you might get a lyric like this:

Ridin’ solo, but I’m never alone,
Got my AI ridin’ shotgun, steady as a stone.

On the endless road, under the sky so vast,
A ghost driver at the wheel, footprints of the past.

Ask it to write a song about self-driving cars in the style of Giuseppe Verdi, and ChatGPT sets the scene of a tenor singing,

Ecco la carrozza senza cavallo! (Behold the carriage without a horse!)
Scivola nella notte, silenziosa come il fato. (It glides through the night, silent as fate.)
L'ingegno dell'uomo, l'orgoglio del progresso, (Man's ingenuity, the pride of progress.)
In questa danza d'acciaio, tessiamo il nostro destino. (In this steel dance, we weave our destiny.)

Though unlikely to win many awards, at least so far, chatbots such as ChatGPT make our smart devices sound stupid. They not only know foreign languages, but also coding languages; they can quickly summarize lengthy legal and financial documents; they are starting to diagnose medical conditions; they can pass the bar exam without studying. On the flip side, we may get tricked into thinking that A.I. models are actually—rather than artificially—intelligent, and that they understand the meaning and implications of the content they deliver. They do not. They are, in the words of the linguist Emily Bender and three co-authors, “stochastic parrots.” It shouldn't be forgotten that, before A.I. could be considered intelligent, it had to swallow up a vast tranche of human intelligence. And, before we learned how to collaborate with robots, robots had to be taught how to collaborate with us.

To even begin to understand how these chatbots work, we had to master new vocabulary, from “large language models” (L.L.M.s) and “neural networks” to “natural-language processing” (N.L.P.) and “generative A.I.” By now, we know the broad strokes: chatbots gobbled up the Internet and analyzed it with a kind of machine learning that mimics the human brain; they string together words statistically, based on which words and phrases typically belong together. Still, the sheer inventiveness of artificial intelligence remains largely inscrutable, as we found out when chatbots “hallucinate.”

Google's Bard, for example, invented information about the James Webb telescope. Microsoft's Bing insisted that the singer Billie Eilish performed at the 2023 Super Bowl halftime show. “I did not comprehend that ChatGPT could fabricate cases,” said an attorney whose federal court brief was found to be full of phony citations and made-up judicial opinions supplied by ChatGPT. (The court issued a fine of five thousand dollars.) In fine print, ChatGPT acknowledges that it may not be reliable: “ChatGPT can make mistakes. Consider checking important information.” Weirdly, a recent study suggests that, in the last year, ChatGPT has grown less accurate when asked to perform certain tasks. Researchers theorize that this has something to do with the material that it's trained on—but, since OpenAI won't share what it is using to train its L.L.M., this is just conjecture.

The knowledge that chatbots make mistakes has not stopped high-school and college students from being some of their most avid early adopters, using chatbots to research and write their papers, complete problem sets, and write code. (During finals week, last May, a student of mine took a walk through the library and saw that just about every laptop was open to ChatGPT.) More than half of young people who responded to a recent Junior Achievement survey said that using a chatbot to help with schoolwork was, in their view, cheating. Yet nearly half said that they were likely to use it.

School administrators were no less conflicted. They couldn't seem to decide if chatbots are agents of deception or tools for learning. In January, David Banks, the New York City schools chancellor, banned

ChatGPT; a spokesperson told the *Washington Post* that the chatbot “does not build critical-thinking and problem-solving skills, which are essential for academic and lifelong success.” Four months later, Banks reversed the ban, calling it “knee-jerk” and fear-based, and saying that it “overlooked the potential of generative AI to support students and teachers, as well as the reality that our students are participating in and will work in a world where understanding generative AI is crucial.” Then there was a professor at Texas A&M who decided to use ChatGPT to root out students who cheated with ChatGPT. After the bot determined that the whole class had done so, the professor threatened to fail everyone. The problem was that ChatGPT was hallucinating. (There are other A.I. programs to catch cheaters; chatbot detection is a growth industry.) In a sense, we are all that professor, beta-testing products whose capacities we may overestimate, misconstrue, or simply not understand.

Artificial intelligence is already used to generate financial reports, ad copy, and sports news. In March, Greg Brockman, a co-founder of OpenAI and its president, predicted—cheerfully—that in the future chatbots would also help write film scripts, and rewrite scenes that viewers didn’t like. Two months later, the Writers Guild of America went on strike, demanding a contract that would protect us all from crummy A.I.-generated movies. They sensed that any A.I. platform that is able to produce credible work in many human domains could be an existential threat to creativity itself.

In September, while screenwriters were negotiating an end to their five-month strike, having persuaded the studios to swear off A.I. scripts, the Authors Guild, along with a group of prominent novelists, filed a class-action suit against OpenAI. They alleged that, when the company vacuumed up the Web, it used their copyrighted work without consent or compensation. Though the writers couldn’t know for sure that the company had appropriated their books, given OpenAI’s less-than-open policy on sharing its training data, the complaint noted that, early on, ChatGPT would respond to queries about specific books with verbatim quotations, “suggesting that the underlying LLM must have ingested these books in their entirety.” (Now the chatbot has been retrained to say, “I can’t provide verbatim excerpts from copyrighted texts.”) Some businesses now sell prompts to help users to impersonate well-known writers. And a writer who can be effortlessly impersonated might not be worth very much.

In July, a report from the literary nonprofit PEN America said that generative A.I. threatens free expression by “supercharging” the dissemination of disinformation and online abuse. “There is also the potential for people to lose trust in language itself, and thus in one another,” the report pointed out. These dangers now extend beyond the written word. OpenAI unveiled an engine, DALL-E 2, that turns text into artificial images; a few months later, Stability AI released a similar tool, Stable Diffusion. According to the Center for Artistic Inquiry and Reporting, A.I.-generated art is “vampirical, feasting on past generations of artwork,” and arguably amounts to “the greatest art heist in history.” As fun and magical as it is to “make” “art” in this way, especially for those of us who are artistically challenged, photorealistic scenes of events that haven’t happened are a threat to the truth, too. Anyone can prompt an A.I. agent to produce an image of a man stuffing a ballot box, or of a confrontation between protesters and police. (I know, because I did, and the results were somewhat realistic.)

While efforts are under way to watermark A.I.-generated images, so far, researchers have not been able to develop a watermarking system that can’t be undermined by widely available tools; they have also been able to add fake watermarks to real images. OpenAI still allows users to remove watermarks at will.

In March, more than a thousand technologists, including [Elon Musk](#) and Steve Wozniak, a co-founder of Apple, signed a letter calling on A.I. companies to pause work on their most advanced technology for six months, to make room for some kind of regulation. It read, in part:

Should we let machines flood our information channels with propaganda and untruth? Should we automate away all the jobs, including the fulfilling ones? Should we develop nonhuman minds that might eventually outnumber, outsmart, obsolete and replace us? Should we risk loss of control of our civilization? Such decisions must not be delegated to unelected tech leaders.

These were not theoretical concerns. A research team from I.B.M., for example, needed only five minutes to trick ChatGPT into writing highly persuasive phishing e-mails. Other researchers have used generative A.I. to write malware that can bypass safety protocols, making it a potential resource for cybercriminals. Goldman Sachs has estimated that A.I. could soon replace three hundred million full-time jobs.

Not surprisingly, there was no pause, and there has been no meaningful regulation. Instead, at the end of October, the Biden Administration issued an “Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence,” a lengthy document that reads more like a wish list than an order. It suggests that the executive branch is engaged in a complicated balancing act between A.I.’s perils and possibilities. Just a week later, OpenAI announced a new suite of products, including an A.I. model that can read prompts as long as a three-hundred-page book; a build-your-own-chatbot kit; and something called “copyright shield,” which promises to pay the legal fees of developers accused of copyright infringement.

With the help of these new tools, I was able to use ChatGPT to create a chatbot that determines which medications are not safe to take together, and another that lists all the restaurants in a particular location that can accommodate specific food allergies and prohibitions. Making these chatbots was intuitive and simple, yet I remained ignorant of the algorithms driving them, the provenance of their training data—was I somehow in breach of copyright?—and whether the information produced by the chatbots was accurate. I also had no clue how much computing power I was using, or what my environmental impact might be. But, hey, they were cool, and the kind of thing people might pay for.

The commercial development of generative A.I. is likely to continue unabated. A.I. will influence an increasing number of complex activities, such as radiology, drug discovery, psychotherapy, hiring, and college admissions. Companies will build it into the next generation of hardware. Samsung, for example, is likely to incorporate generative A.I. into its forthcoming flagship phones, which it will unveil in January. [Sam Altman](#), the OpenAI co-founder who recently boomeranged out of—and then back into—the role of C.E.O., has reportedly been collaborating with Jony Ive, the famed Apple designer, to create “the iPhone of artificial intelligence.” We may look back on 2023 with a kind of nostalgia, for a time when intelligence had not yet become a product. ♦

*[Sue Halpern](#) is a staff writer at *The New Yorker*. She is the author of, most recently, the novel “[Summer Hours at the Robbers Library](#).”*