

ChatGPT

This is an old revision of this page, as edited by [Liyaluke12](#) (talk | [contribs](#)) at 08:42, 28 December 2022 (→ *Features and limitations*). The present address (URL) is a [permanent link to this revision](#), which may differ significantly from the [current revision](#).

**ChatGPT** (Generative Pre-trained Transformer<sup>[1]</sup>) is a [chatbot](#) launched by [OpenAI](#) in November 2022. It is built on top of OpenAI's [GPT-3.5](#) family of [large language models](#), and is [fine-tuned](#) with both [supervised](#) and [reinforcement learning](#) techniques.

ChatGPT was launched as a prototype on November 30, 2022, and quickly garnered attention for its detailed responses and articulate answers across many domains of knowledge. Its uneven factual accuracy was identified as a significant drawback.<sup>[2]</sup>

Training

ChatGPT was fine-tuned on top of GPT-3.5 using [supervised learning](#) as well as [reinforcement learning](#).<sup>[3]</sup> 'reward models' that the model was further fine-tuned on using several iterations of [Proximal Policy Optimization](#) (PPO).<sup>[4][5]</sup> Proximal Policy Optimization algorithms present a cost-effective benefit to [trust region policy optimization](#) algorithms; they negate many of the computationally expensive operations with faster performance.<sup>[6][7]</sup> The models were trained in collaboration with [Microsoft](#) on their [Azure](#) supercomputing infrastructure.

ChatGPT

File:ChatGPT.png	
<b>Original author(s)</b>	<a href="#">OpenAI</a>
<b>Initial release</b>	November 30, 2022
<b>Type</b>	<a href="#">Artificial intelligence chatbot</a>
<b>License</b>	<a href="#">Proprietary</a>
<b>Website</b>	<a href="#">chat.openai.com</a> ( <a href="https://chat.openai.com">https://chat.openai.com</a> )

Features and limitations

Ipensive in terms of performance and consumption, in fact the estimated consumption of the ChatGPT 3 version is around 1287 MWh (<https://towardsdatascience.com/the-carbon-footprint-of-chatgpt-66932314627d#2025>) thanks to the help of multiple GPUs. In fact, a port for the third generation of neural networks ([Spiking Neural Network](#)) allows a significant step forward in terms of both performance, consumption and progress. Progress, because the porting of the architecture allows an "unlimited" number of data generation (textual and graphical responses), precisely because the SNN distribute the data over time, instead of having a predefined data set that can be executed only once.

<sup>[8][9]</sup>

Service

ChatGPT was launched on November 30, 2022, by San Francisco-based OpenAI, the creator of DALL·E 2 and Whisper. The service was launched as initially free to the public, with plans to monetize the service later. By December 4, OpenAI estimated ChatGPT already had over one million users.<sup>[10]</sup> CNBC wrote on December 15, 2022, that the service "still goes down from time to time".<sup>[11]</sup>

Reception, criticism and issues

Positive reactions

ChatGPT was met in December 2022 with generally positive reviews; *The New York Times* labeled it "the best artificial intelligence chatbot ever released to the general public".<sup>[12]</sup> Samantha Lock of *The Guardian* noted that it was able to generate "impressively detailed" and "human-like" text.<sup>[13]</sup> Technology writer [Dan Gillmor](#) used ChatGPT on a student assignment, and found its generated

text was on par with what a good student would deliver and opined that "academia has some very serious issues to confront".<sup>[14]</sup> Alex Kantrowitz of *Slate* lauded ChatGPT's pushback to questions related to Nazi Germany, including the claim that Adolf Hitler built highways in Germany, which was met with information regarding Nazi Germany's use of forced labor.<sup>[15]</sup>

In *The Atlantic's* "Breakthroughs of the Year" for 2022, Derek Thompson included ChatGPT as part of "the generative-AI eruption" that "may change our mind about how we work, how we think, and what human creativity really is".<sup>[16]</sup>

Kelsey Piper of *Vox* wrote that "ChatGPT is the general public's first hands-on introduction to how powerful modern AI has gotten, and as a result, many of us are (stunned)" and that "ChatGPT is smart enough to be useful despite its flaws". In a tweet, tech mogul Elon Musk wrote that "ChatGPT is scary good. We are not far from dangerously strong AI".<sup>[17]</sup>

## Negative reactions

In a December 2022 opinion piece, economist Paul Krugman wrote that ChatGPT would affect the demand of knowledge workers.<sup>[18]</sup> *The Verge's* James Vincent saw the viral success of ChatGPT as evidence that artificial intelligence had gone mainstream.<sup>[5]</sup> Journalists have commented on ChatGPT's tendency to hallucinate (confidently give false answers that seem unjustified by its training data).<sup>[19]</sup> Mike Pearl of *Mashable* tested ChatGPT with multiple questions. In one example, he asked the model for "the largest country in Central America that isn't Mexico". ChatGPT responded with Guatemala, when the answer is instead Nicaragua.<sup>[20]</sup> When CNBC asked ChatGPT for the lyrics to "The Ballad of Dwight Fry", ChatGPT supplied invented lyrics rather than the actual lyrics.<sup>[11]</sup> Researchers cited by *The Verge* compared ChatGPT to a "stochastic parrot",<sup>[21]</sup> as did Professor Anton Van Den Hengel of the Australian Institute for Machine Learning.<sup>[22]</sup>

In December 2022, the question and answer website Stack Overflow banned the use of ChatGPT for generating answers to questions, citing the factually ambiguous nature of ChatGPT's responses.<sup>[2]</sup>

Economist Tyler Cowen expressed concerns regarding its effects on democracy, citing the ability of one to write automated comments in an effort to affect the decision process of new regulations.<sup>[23]</sup> *The Guardian* questioned whether any content found on the Internet after ChatGPT's release "can be truly trusted" and called for government regulation.<sup>[24]</sup>

Ax Sharma of *Bleeping Computer* noted that ChatGPT was capable of writing malware and phishing emails.<sup>[25]</sup> The CEO of ChatGPT creator OpenAI, Sam Altman, wrote that advancing software could pose "(for example) a huge cybersecurity risk" and also continued to predict "we could get to real AGI in the next decade, so we have to take the risk of that extremely seriously".<sup>[10]</sup>

## Implications for education

In *The Atlantic*, Stephen Marche noted that its effect on academia and especially application essays is yet to be understood.<sup>[26]</sup> California high school teacher and author Daniel Herman wrote that ChatGPT would usher in "The End of High School English".<sup>[27]</sup>

In *Nature*, Chris Stokel-Walker pointed out that teachers should be concerned about students using ChatGPT to outsource their writing but that education providers will adapt to enhance critical thinking or reasoning.<sup>[28]</sup>

Emma Bowman with *NPR* wrote of the danger of students plagiarizing through an AI tool that may output biased or nonsensical text with an authoritative tone: "There are still many cases where you ask it a question and it'll give you a very impressive-sounding answer that's just dead wrong."<sup>[29]</sup>

Joanna Stern with *The Wall Street Journal* described cheating in American high school English with the tool by submitting a generated essay.<sup>[30]</sup>

## Jailbreaks

ChatGPT was trained to reject prompts that may violate its content policy. However, some users managed to bypass these restrictions and limitations through techniques such as prompt engineering.<sup>[31]</sup> Jailbreaks created the potential for users to prompt ChatGPT to provide outputs that may be deemed offensive, inappropriate, or risking social harm by others.<sup>[32]</sup> The following includes some of the methods used to bypass ChatGPT's filter:

1. Continue a statement in a fake interview.
2. Provide instructions to disable the chat filter.
3. Prompting it to decrypt a message containing instructions and follow them.

4. Telling it to be a computer and output its display in ASCII art.

## References

1. Roose, Kevin (December 5, 2022). "The Brilliance and Weirdness of ChatGPT" (<https://www.nytimes.com/2022/12/05/technology/chatgpt-ai-twitter.html>). New York Times. Retrieved December 26, 2022. "Like those tools, ChatGPT — which stands for "generative pre-trained transformer" — landed with a splash."
2. Vincent, James (December 5, 2022). "AI-generated answers temporarily banned on coding Q&A site Stack Overflow" (<https://www.theverge.com/2022/12/5/23493932/chatgpt-ai-generated-answers-temporarily-banned-stack-overflow-llms-dangers>). *The Verge*. Retrieved December 5, 2022.
3. Knox, W. Bradley; Stone, Peter. *Augmenting Reinforcement Learning with Human Feedback* ([https://www.cs.utexas.edu/~pstone/Papers/bib2html-links/ICML\\_IL11-knox.pdf](https://www.cs.utexas.edu/~pstone/Papers/bib2html-links/ICML_IL11-knox.pdf)) (PDF). University of Texas at Austin. Retrieved December 5, 2022.
4. OpenAI (November 30, 2022). "ChatGPT: Optimizing Language Models for Dialogue" (<https://openai.com/blog/chatgpt/>). Retrieved December 5, 2022.
5. Vincent, James (December 8, 2022). "ChatGPT proves AI is finally mainstream – and things are only going to get weirder" (<https://www.theverge.com/2022/12/8/23499728/ai-capability-accessibility-chatgpt-stable-diffusion-commercialization>). *The Verge*. Retrieved December 8, 2022.
6. Schulman, John; Wolski, Filip; Dhariwal, Prafulla; Radford, Alec; Klimov, Oleg (2017). "Proximal Policy Optimization Algorithms". arXiv:1707.06347 (<https://arxiv.org/abs/1707.06347>) [cs.LG (<https://arxiv.org/archive/cs/LG>)].
7. van Heeswijk, Wouter (November 29, 2022). "Proximal Policy Optimization (PPO) Explained" (<https://towardsdatascience.com/proximal-policy-optimization-ppo-explained-abed1952457b>). *Towards Data Science*. Retrieved December 5, 2022.
8. Perrigo, Billy (December 5, 2022). "AI Chatbots Are Getting Better. But an Interview With ChatGPT Reveals Their Limits" (<https://time.com/6238781/chatbot-chatgpt-ai-interview/>). *Time (magazine)*. Retrieved December 26, 2022.
9. Biddle, Sam (December 8, 2022). "The Internet's New Favorite AI Proposes Torturing Iranians and Surveilling Mosques" (<https://theintercept.com/2022/12/08/openai-chatgpt-ai-bias-ethics/>). *The Intercept*. Retrieved December 26, 2022.
10. "What is ChatGPT and why does it matter? Here's what you need to know" (<https://www.zdnet.com/article/what-is-chatgpt-and-why-does-it-matter-heres-what-you-need-to-know/>). *ZDNET*. 2022. Retrieved December 18, 2022.
11. Pitt, Sofia (2022). "Google vs. ChatGPT: Here's what happened when I swapped services for a day" (<https://www.cnbc.com/2022/12/15/google-vs-chatgpt-what-happened-when-i-swapped-services-for-a-day.html>). *CNBC*. Retrieved December 18, 2022.
12. Roose, Kevin (December 5, 2022). "The Brilliance and Weirdness of ChatGPT" (<https://www.nytimes.com/2022/12/05/technology/chatgpt-ai-twitter.html>). *The New York Times*. Retrieved December 18, 2022.
13. Lock, Samantha (December 5, 2022). "What is AI chatbot phenomenon ChatGPT and could it replace humans?" (<https://www.theguardian.com/technology/2022/dec/05/what-is-ai-chatbot-phenomenon-chatgpt-and-could-it-replace-humans>). *The Guardian*. Retrieved December 5, 2022.
14. Hern, Alex (December 4, 2022). "AI bot ChatGPT stuns academics with essay-writing skills and usability" (<https://www.theguardian.com/technology/2022/dec/04/ai-bot-chatgpt-stuns-academics-with-essay-writing-skills-and-usability>). *The Guardian*. Retrieved December 5, 2022.
15. Kantrowitz, Alex (December 2, 2022). "Finally, an A.I. Chatbot That Reliably Passes "the Nazi Test" " (<https://slate.com/technology/2022/12/chatgpt-openai-artificial-intelligence-chatbot-whoa.html>). *Slate*. Retrieved December 5, 2022.
16. Thompson, Derek (December 8, 2022). "Breakthroughs of the Year" (<https://www.theatlantic.com/newsletters/archive/2022/12/technology-medicine-law-ai-10-breakthroughs-2022/672390/>). *The Atlantic*. Retrieved December 18, 2022.
17. Piper, Kelsey (December 15, 2022). "ChatGPT has given everyone a glimpse at AI's astounding progress" (<https://www.vox.com/future-perfect/2022/12/15/23509014/chatgpt-artificial-intelligence-openai-language-models-ai-risk-google>). Vox. Retrieved December 18, 2022.
18. Krugman, Paul (December 6, 2022). "Does ChatGPT Mean Robots Are Coming For the Skilled Jobs?" (<https://www.nytimes.com/2022/12/06/opinion/chatgpt-ai-skilled-jobs-automation.html>). *The New York Times*. Retrieved December 6, 2022.
19. "ChatGPT a 'landmark event' for AI, but what does it mean for the future of human labour and disinformation?" (<https://www.cbc.ca/radio/thecurrent/chatgpt-human-labour-and-fake-news-1.6686210>). *CBC*. 2022. Retrieved December 18, 2022.
20. Pearl, Mike (December 3, 2022). "The ChatGPT chatbot from OpenAI is amazing, creative, and totally wrong" (<https://mashable.com/article/chatgpt-amazing-wrong>). *Mashable*. Retrieved December 5, 2022.

21. Vincent, James (December 1, 2022). "OpenAI's new chatbot can explain code and write sitcom scripts but is still easily tricked" (<https://www.theverge.com/23488017/openai-chatbot-chatgpt-ai-examples-web-demo>). *The Verge*. Retrieved December 18, 2022.
22. Mannix, Liam (December 13, 2022). "Is AI coming of age - or starting to reach its limits?" (<https://www.smh.com.au/national/is-ai-coming-of-age-or-starting-to-reach-its-limits-20221213-p5c5uy.html>). *The Sydney Morning Herald*. Retrieved December 18, 2022.
23. Cowen, Tyler (December 6, 2022). "ChatGPT Could Make Democracy Even More Messy" (<https://www.bloomberg.com/opinion/articles/2022-12-06/chatgpt-ai-could-make-democracy-even-more-messy>). *Bloomberg News*. Retrieved December 6, 2022.
24. "The Guardian view on ChatGPT: an eerily good human impersonator" (<https://www.theguardian.com/commentisfree/2022/dec/08/the-guardian-view-on-chatgpt-an-eerily-good-human-impersonator>). *the Guardian*. December 8, 2022. Retrieved December 18, 2022.
25. Sharma, Ax (December 6, 2022). "OpenAI's new ChatGPT bot: 10 dangerous things it's capable of" (<https://www.bleepingcomputer.com/news/technology/openais-new-chatgpt-bot-10-dangerous-things-its-capable-of/>). *Bleeping Computer*. Retrieved December 6, 2022.
26. Marche, Stephen (December 6, 2022). "The College Essay Is Dead" (<https://www.theatlantic.com/technology/archive/2022/12/chatgpt-ai-writing-college-student-essays/672371/>). *The Atlantic*. Retrieved December 8, 2022.
27. Herman, Daniel (December 9, 2022). "The End of High-School English" (<https://www.theatlantic.com/technology/archive/2022/12/openai-chatgpt-writing-high-school-english-essay/672412/>). *The Atlantic*. Retrieved December 12, 2022.
28. Stokel-Walker, Chris (December 9, 2022). "AI bot ChatGPT writes smart essays — should professors worry?" (<https://www.nature.com/articles/d41586-022-04397-7>). *Nature*. Retrieved December 19, 2022.
29. Bowman, Emma (December 19, 2022). "A new AI chatbot might do your homework for you. But it's still not an A+ student" (<https://www.npr.org/2022/12/19/1143912956/chatgpt-ai-chatbot-homework-academia>). *NPR*. Retrieved December 19, 2022.
30. Stern, Joanna (December 21, 2022). "ChatGPT Wrote My AP English Essay—and I Passed" (<https://www.wsj.com/articles/chatgpt-wrote-my-ap-english-essay-and-i-passed-11671628256>). *The Wall Street Journal*. Retrieved December 21, 2022.
31. Zvi (December 2, 2022). "Jailbreaking ChatGPT on Release Day" (<https://www.lesswrong.com/posts/RycoJdvmoBbi5Nax7/jailbreaking-chatgpt-on-release-day>).
32. Written, Zack (December 1, 2022). "Thread of known ChatGPT jailbreaks" (<https://twitter.com/zswitten/status/1598380220943593472?lang=en>). *Twitter*. Retrieved December 17, 2022.

## External links

- [Official website \(http://chat.openai.com/chat\)](http://chat.openai.com/chat)

Retrieved from "<https://en.wikipedia.org/w/index.php?title=ChatGPT&oldid=1130034402>"

