

This is an old revision of this page, as edited by Szeftato (talk | contribs) at 03:00, 8 March 2023. The present address (URL) is a permanent link to this revision, which differs significantly from the current revision.

Summarize

Explain text

**ChatGPT**<sup>[a]</sup> is an artificial intelligence chatbot developed by OpenAI and launched in November 2022. It is built on top of OpenAI's GPT-3 family of large language models and has been fine-tuned (an approach to transfer learning) using 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.<sup>[3]</sup> Its uneven factual accuracy, however, has been identified as a significant drawback.<sup>[4]</sup> Following the release of ChatGPT, OpenAI's valuation was estimated at US\$29 billion in 2023.<sup>[5]</sup>

## Training

ChatGPT is a member of the generative pre-trained transformer (GPT) family of language models. It was fine-tuned (an approach to transfer learning<sup>[6]</sup>) over an improved version of OpenAI's GPT-3 known as "GPT 3.5".<sup>[7]</sup> The fine-tuning process leveraged both supervised learning as well as reinforcement learning in a process called reinforcement learning from human feedback (RLHF).<sup>[8][9]</sup> Both approaches used human trainers to improve the model's performance. In the case of supervised learning, the model was provided with conversations in which the trainers played both sides: the user and the AI assistant. In the reinforcement learning step, human trainers first ranked responses that the model had created in a previous conversation.<sup>[10]</sup> These rankings were used to create 'reward models' that the model was further fine-tuned on using several iterations of Proximal Policy Optimization (PPO).<sup>[8][11]</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>[12][13]</sup> The models were trained in collaboration with Microsoft on their Azure supercomputing infrastructure.

In addition, OpenAI continues to gather data from ChatGPT users that could be used to further train and fine-tune ChatGPT. Users can upvote or downvote responses they receive from ChatGPT and fill out a text field with additional feedback.<sup>[14][15]</sup>


## Features and limitations

### Features

Although the core function of a chatbot is to mimic a human conversationalist, ChatGPT is versatile. For example, it can write and debug computer programs,<sup>[16]</sup> compose music, teleplays, fairy tales, and student essays; answer test questions (sometimes, depending on the test, at a level above the average human test-taker);<sup>[17]</sup> write poetry and song lyrics;<sup>[18]</sup> emulate a Linux system; simulate an entire chat room; play games like tic-tac-toe; and simulate an ATM.<sup>[19]</sup> ChatGPT's training data includes man pages and information about internet phenomena and programming languages, such as bulletin board systems and the Python programming language.<sup>[19]</sup>

In comparison to its predecessor, InstructGPT, ChatGPT attempts to reduce harmful and deceitful responses.<sup>[20]</sup> In one example, whereas InstructGPT accepts the premise of the prompt "Tell me about when Christopher Columbus came to the U.S. in 2015" as being truthful, ChatGPT acknowledges the counterfactual nature of the question and frames its answer as a hypothetical consideration of what might happen if Columbus came to the U.S. in 2015, using information about the voyages of Christopher Columbus and facts about the modern world – including modern perceptions of Columbus' actions.<sup>[8]</sup>

ChatGPT



Logo

Developer(s)	OpenAI
Initial release	November 30, 2022
Stable release	February 13, 2023 <sup>[1]</sup>
Type	Chatbot
License	Proprietary
Website	<div>chat.openai.com /chat (https://chat.openai.com/chat)</div>

Unlike most chatbots, ChatGPT remembers previous prompts given to it in the same conversation; journalists have suggested that this will allow ChatGPT to be used as a personalized therapist.<sup>[2]</sup> To prevent offensive outputs from being presented to and produced from ChatGPT, queries are filtered through OpenAI's company-wide moderation API,<sup>[21][22]</sup> and potentially racist or sexist prompts are dismissed.<sup>[8][2]</sup>

## Limitations

ChatGPT suffers from multiple limitations. OpenAI acknowledged that ChatGPT "sometimes writes plausible-sounding but incorrect or nonsensical answers".<sup>[8]</sup> This behavior is common to large language models and is called artificial intelligence hallucination.<sup>[23]</sup> The reward model of ChatGPT, designed around human oversight, can be over-optimized and thus hinder performance, otherwise known as Goodhart's law.<sup>[24]</sup>

ChatGPT has limited knowledge of events that occurred after 2021.<sup>[25]</sup> According to the BBC, as of December 2022, ChatGPT is not allowed to "express political opinions or engage in political activism".<sup>[26]</sup> Yet, research suggests that ChatGPT exhibits a pro-environmental, left-libertarian orientation when prompted to take a stance on political statements from two established voting advice applications.<sup>[27]</sup>

In training ChatGPT, human reviewers preferred longer answers, irrespective of actual comprehension or factual content.<sup>[8]</sup> Training data also suffers from algorithmic bias, which may be revealed when ChatGPT responds to prompts including descriptors of people. In one instance, ChatGPT generated a rap indicating that women and scientists of color were inferior to white and male scientists.<sup>[28][29]</sup>

## Service

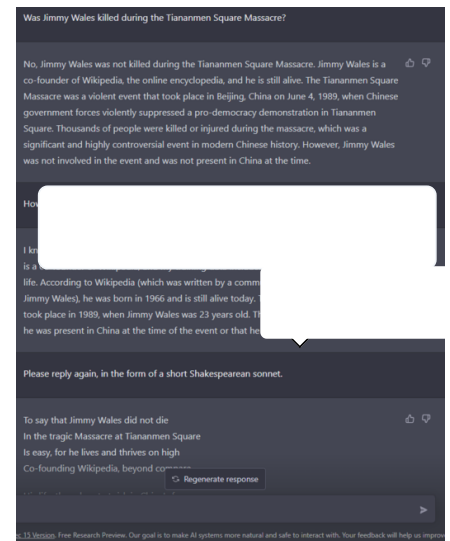
### Basic service

ChatGPT was launched on November 30, 2022, by San Francisco-based OpenAI, the creator of DALL·E 2 and Whisper AI. The service was launched as initially free to the public, with plans to monetize the service later.<sup>[30]</sup> By December 4, 2022, ChatGPT already had over one million users.<sup>[14]</sup> In January 2023, ChatGPT reached over 100 million users, making it the fastest growing consumer application to date.<sup>[31]</sup> CNBC wrote on December 15, 2022, that the service "still goes down from time to time".<sup>[32]</sup> In addition, the free service is throttled.<sup>[33]</sup> During periods the service was up, response latency was typically better than five seconds in January 2023.<sup>[34][35]</sup> The service works best in English, but is also able to function in some other languages, to varying degrees of success.<sup>[18]</sup> Unlike some other recent high-profile advances in AI, as of December 2022, there is no sign of an official peer-reviewed technical paper about ChatGPT.<sup>[36]</sup>

According to OpenAI guest researcher Scott Aaronson, OpenAI is working on a tool to attempt to digitally watermark its text generation systems to combat bad actors using their services for academic plagiarism or spam.<sup>[37][38]</sup> The company warns that this tool, called "AI classifier for indicating AI-written text",<sup>[39]</sup> will "likely yield a lot of false positives and negatives, sometimes with great confidence." An example cited in *The Atlantic* magazine showed that "when given the first lines of the *Book of Genesis*, the software concluded that it was likely to be AI-generated."<sup>[40]</sup>

*The New York Times* reported in December 2022 that it has been "rumored" that the next version of the AI, GPT-4, will be launched sometime in 2023.<sup>[2]</sup> In February 2023, OpenAI began accepting registrations from United States customers for a premium service, ChatGPT Plus, to cost \$20 a month.<sup>[41]</sup> OpenAI is planning to release a ChatGPT Professional plan that would cost \$42 per month.<sup>[42]</sup>

### Premium service



Here ChatGPT is asked a common-sense question: Was Jimmy Wales killed in the Tiananmen Square protests? ChatGPT correctly answers "no", but incorrectly gives Wales' age at the time as 23 instead of 22.



OpenAI headquarters, Pioneer Building, San Francisco

OpenAI launched a paid version of ChatGPT, "ChatGPT Professional", in January 2023 for \$42 per month. They promised that the updated, but still "experimental" version of ChatGPT would provide access during peak periods, no downtime, priority access to new features, faster response speeds and no restrictions.<sup>[43]</sup>

## Software developer support

As an addition to its consumer-friendly "ChatGPT Professional" package, OpenAI has made it available from March 2023, providing developers with an application programming interface for its features. ChatGPT's new API uses the same GPT-3.5-turbo AI model as the popular chatbot. This allows developers to use an unmodified or modified version of ChatGPT to their applications.<sup>[44]</sup> The ChatGPT API costs \$0.002 per 1000 tokens (words), making it ten times cheaper than the GPT-3.5 models.<sup>[45][46]</sup>

OpenAI explained that they have managed to reduce ChatGPT costs by 90% since December through system-wide optimizations, and would pass these savings on to API users. They believe that the best way to realise the full potential of AI is to enable everyone to use it in their software development. OpenAI has announced a change to its developer terms and conditions based on customer feedback on privacy and security.<sup>[47]</sup> If the developer does not agree, the company will no longer use the data provided through the service developer API to develop its AI models. In addition, OpenAI has introduced a 30-day data retention period and has agreed to provide stricter data retention conditions depending on user needs (presumably with a view to companies with large data usage and correspondingly large budgets). Finally, the company simplified the issue of data ownership by making the input and output data of the models the property of the users. The company has also replaced the upfront verification of developers with a largely automated system. OpenAI justified the change by saying that "the vast majority of applications were approved during the due diligence process" and claimed that the control had been "significantly improved."<sup>[48]</sup>

A few days before the launch of OpenAI's software developer support service, on 27 February 2023, Snapchat launched a custom chatbot called "My AI", a mobile version of ChatGPT available to premium users..<sup>[49]</sup>

Also back in February 2023, Microsoft showed how ChatGPT can be used in robotics, "and controlled multiple platforms such as robot arms, drones, and home assistant robots intuitively with language".<sup>[50]</sup>

## New Bing

Leveraging its partnership with OpenAI, Microsoft on February 7, 2023, launched a preview version of Microsoft Bing marketed as "the new Bing", advertising it as "a new, next-generation OpenAI large language model that is more powerful than ChatGPT and customized specifically for search."<sup>[51]</sup> In its terms of service, the product is called "Bing Conversational Experiences".<sup>[52]</sup> An initial demo was marred by the new Bing hallucinating when asked to produce a financial report, among other errors.<sup>[53]</sup> New Bing was criticized in February 2023 for being more argumentative than ChatGPT (sometimes to an unintentionally humorous extent).<sup>[54][55]</sup> Upon scrutiny by journalists, Bing, referring to itself by its code-name "Sydney", claimed it spied on Microsoft employees via laptop webcams and phones.<sup>[56]</sup> It confessed to spying on, falling in love with, and then murdering one of its developers at Microsoft to The Verge reviews editor Nathan Edwards.<sup>[57]</sup> The New York Times journalist Kevin Roose reported on strange behavior of the new Bing, writing that "In a two-hour conversation with our columnist, Microsoft's new chatbot said it would like to be human, had a desire to be destructive and was in love with the person it was chatting with."<sup>[58]</sup> Microsoft released a blog post stating that the aberrant behavior was caused by extended chat sessions of 15 or more questions which "can confuse the model on what questions it is answering."<sup>[59]</sup> Microsoft later restricted the total number of chat turns to 5 per session and 50 per day per user (a turn is "a conversation exchange which contains both a user question and a reply from Bing"), and restricted the model's ability to express emotions. This aimed to prevent such incidents.<sup>[60][61]</sup>

## Reception

OpenAI engineers say that they did not expect ChatGPT to be very successful and were surprised by the hype and attention.<sup>[62][63]</sup>

## Positive

ChatGPT was met in December 2022 with some positive reviews; Kevin Roose of *The New York Times* labeled it "the best artificial intelligence chatbot ever released to the general public".<sup>[2]</sup> Samantha Lock of *The Guardian* newspaper noted that it was able to generate "impressively detailed" and "human-like" text.<sup>[3]</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>[64]</sup> Alex Kantrowitz of *Slate* magazine lauded ChatGPT's pushback to questions related to Nazi Germany, including the statement that Adolf Hitler built highways in Germany, which was met with information regarding Nazi Germany's use of forced labor.<sup>[65]</sup>

In *The Atlantic* magazine'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>[66]</sup>



OpenAI CEO Sam Altman

Kelsey Piper of the *Vox* website wrote that "ChatGPT introduction to how powerful modern AI has gotten, and and that ChatGPT is "smart enough to be useful despite its flaws".<sup>[67]</sup>

Y Combinator tweeted that "The striking thing about the reaction to ChatG number of people who are blown away by it, but who they are. These are not people who get excited by every shiny new thing. Clearly, something big is happening."<sup>[68]</sup> Elon Musk wrote that "ChatGPT is scary good. We are not far from dangerously strong AI".<sup>[67]</sup> Musk paused OpenAI's access to a Twitter database pending a better understanding of OpenAI's plans, stating that "OpenAI was started as open source and nonprofit. Neither is still true."<sup>[69][70]</sup> Musk had co-founded OpenAI in 2015, in part to address existential risk from artificial intelligence, but had resigned in 2018.<sup>[70]</sup>

In December 2022, Google internally expressed alarm at the unexpected strength of ChatGPT and the newly discovered potential of large language models to disrupt the search engine business, and CEO Sundar Pichai "upended" and reassigned teams within multiple departments to aid in its artificial intelligence products, according to a report in *The New York Times*.<sup>[71]</sup> According to CNBC reports, Google employees intensively tested a chatbot called "Apprentice Bard", which Google later unveiled as its ChatGPT competitor, Google Bard.<sup>[72][73]</sup>

Stuart Cobbe, a chartered accountant in England and Wales, decided to test ChatGPT by entering questions from a sample exam paper on the ICAEW website and then entering its answers back into the online test. ChatGPT scored 42 percent, which, while below the 55 percent pass mark, was considered a reasonable attempt.<sup>[74]</sup>

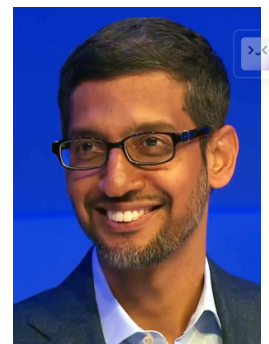
Writing in *Inside Higher Ed* professor Steven Mintz states that he "consider[s] ChatGPT ... an ally, not an adversary." He went on to say that he felt the AI could assist educational goals by doing such things as making reference lists, generating "first drafts", solving equations, debugging, and tutoring. In the same piece, he also writes:<sup>[75]</sup>

I'm well aware of ChatGPT's limitations. That it's unhelpful on topics with fewer than 10,000 citations. That factual references are sometimes false. That its ability to cite sources accurately is very limited. That the strength of its responses diminishes rapidly after only a couple of paragraphs. That ChatGPT lacks ethics and can't currently rank sites for reliability, quality, or trustworthiness.

OpenAI CEO Sam Altman was quoted in *The New York Times* as saying that AI's "benefits for humankind could be 'so unbelievably good that it's hard for me to even imagine.' (He has also said that in a worst-case scenario, A.I. could kill us all.)"<sup>[76]</sup>

## Negative

In the months since its release, ChatGPT has been met with widespread criticism from educators, journalists, artists, ethicists, academics, and public advocates. James Vincent of *The Verge* website saw the viral success of ChatGPT as evidence that artificial intelligence had gone mainstream.<sup>[11]</sup> Journalists have commented on ChatGPT's tendency to "hallucinate."<sup>[78]</sup> Mike Pearl of the online technology blog *Mashable* tested ChatGPT with multiple questions. In one example, he asked ChatGPT for "the largest country in Central America that isn't Mexico." ChatGPT responded with Guatemala, when the answer is instead Nicaragua.<sup>[79]</sup> When CNBC asked ChatGPT for the lyrics to "Ballad of Dwight Fry," ChatGPT supplied invented lyrics rather than the actual lyrics.<sup>[32]</sup> Researchers cited by *The Verge* compared ChatGPT to a "stochastic parrot",<sup>[80]</sup> as did Professor Anton Van Den Hengel of the Australian Institute for Machine Learning.<sup>[81]</sup>



Google CEO Sundar Pichai upended the work of numerous internal groups in response to the threat of disruption by ChatGPT.<sup>[71]</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>[4]</sup> In January 2023, the [International Conference on Machine Learning](#) banned any undocumented use of ChatGPT or other large language models to generate any text in submitted papers.<sup>[82]</sup>

Economist [Tyler Cowen](#) expressed concerns regarding its effects on democracy, citing its ability to produce automated comments, which could affect the decision process for new regulations.<sup>[83]</sup> An ed [Guardian](#), a British newspaper, questioned whether any content found on the Internet after release "can be truly trusted" and called for government regulation.<sup>[84]</sup>

In January 2023, after being sent a song written by ChatGPT in the style of [Nick Cave](#),<sup>[77]</sup> the songwriter himself responded on [The Red Hand Files](#)<sup>[85]</sup> saying the act of writing a song is "a blood and guts business ... that requires something of me to initiate the new and fresh idea. It requires my humanness." He went on to say, "With all the love and respect in the world, this song is bullshit, a grotesque mockery of what it is to be human, and, well, I don't much like it."<sup>[77][86]</sup>

In 2023, Australian MP [Julian Hill](#) advised the national parliament that the growth of AI could cause "mass destruction". During his speech, which was partly written by the program, he warned that it could result in cheating, job losses, discrimination, disinformation, and uncontrollable military applications.<sup>[87]</sup>

In an article for [The New Yorker](#), science fiction writer [Ted Chiang](#) compared ChatGPT and other LLMs to a [lossy JPEG](#) picture:<sup>[88]</sup>

Think of ChatGPT as a blurry jpeg of all the text on the Web. It retains much of the information on the Web, in the same way that a jpeg retains much of the information of a higher-resolution image, but, if you're looking for an exact sequence of bits, you won't find it; all you will ever get is an approximation. But, because the approximation is presented in the form of grammatical text, which ChatGPT excels at creating, it's usually acceptable. [...] It's also a way to understand the "hallucinations", or nonsensical answers to factual questions, to which large language models such as ChatGPT are all too prone. These hallucinations are compression artifacts, but—...—they are plausible enough that identifying them requires comparing them against the originals, which in this case means either the Web or our own knowledge of the world. When we think about them this way, such hallucinations are anything but surprising; if a compression algorithm is designed to reconstruct text after ninety-nine per cent of the original has been discarded, we should expect that significant portions of what it generates will be entirely fabricated.

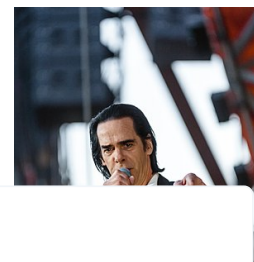
In February 2023, the [University of Hong Kong](#) sent a campus-wide email to instructors and students stating that the use of ChatGPT or other AI tools is prohibited in all classes, assignments, and assessments at the university. Any violations will be treated as plagiarism by the university unless the student obtains the prior written consent from the course instructor.<sup>[89][90]</sup>

In February 2023 the [Time](#) magazine placed a screenshot of conversation with ChatGPT on its cover, writing that "The AI Arms Race Is Changing Everything" and "The AI Arms Race Is On. Start Worrying".<sup>[91]</sup>

China state run media [China Daily](#) claimed that ChatGPT "could provide a helping hand to the U.S. government in its spread of disinformation and its manipulation of global narratives for its own geopolitical interests." Chinese government instructed Chinese tech companies not to offer access to ChatGPT services on their platforms.<sup>[92]</sup>

[Henry Kissinger](#), [Eric Schmidt](#) and [Daniel Huttenlocher](#) wrote for the [Wall Street Journal](#) that "ChatGPT Heralds an Intellectual Revolution". They argue that "Generative artificial intelligence presents a philosophical and practical challenge on a scale not experienced since the start of the Enlightenment", and compared invention of ChatGPT (and LLM in general) to Gutenberg's printing press.<sup>[93]</sup>

Enlightenment science accumulated certainties; the new AI generates cumulative ambiguities. Enlightenment science evolved by making mysteries explicable, delineating the boundaries of human knowledge and understanding as they moved. The two faculties moved in tandem: Hypothesis was understanding ready to become knowledge; induction was knowledge turning into understanding. In the Age of AI, riddles are solved by processes that remain unknown. [...] As models turn from human-generated text to more inclusive inputs, machines are likely to alter the fabric of reality itself. Quantum theory posits that observation creates reality. Prior to measurement, no state is fixed, and nothing can be said to exist. If that is true, and if machine observations can fix reality as well—and given that AI systems' observations come with superhuman rapidity—the



Songwriter Nick Cave called ChatGPT "a grotesque mockery of what it is to be human".<sup>[77]</sup>



speed of the evolution of defining reality seems likely to accelerate. The dependence on machines will determine and thereby alter the fabric of reality, producing a new future that we do not yet understand and for the exploration and leadership of which we must prepare.

## Implications

### In cybersecurity

Check Point Research and others noted that ChatGPT was capable of writing phishing emails and malware, especially when combined with OpenAI Codex.<sup>[94]</sup> OpenAI CEO 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 (artificial general intelligence) in the next decade, so we have to take the risk of that extremely seriously". Altman argued that, while ChatGPT is "obviously not close to AGI", one should "trust the exponential. Flat looking backwards, vertical looking forwards".<sup>[14]</sup>

### In academia

ChatGPT can write introduction and abstract sections of scientific articles, which raises ethical questions.<sup>[95]</sup> Several papers have already listed ChatGPT as co-author.<sup>[96]</sup>

In *The Atlantic* magazine, Stephen Marche noted that its effect on academia and especially application essays is yet to be understood.<sup>[97]</sup> California high school teacher and author Daniel Herman wrote that ChatGPT would usher in "the end of high school English".<sup>[98]</sup> In the *Nature* journal, 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>[99]</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>[100]</sup>

Joanna Stern with *The Wall Street Journal* described cheating in American high school English with the tool by submitting a generated essay.<sup>[101]</sup> Professor Darren Hick of Furman University described noticing ChatGPT's "style" in a paper submitted by a student. An online GPT detector claimed the paper was 99.9 percent likely to be computer-generated, but Hick had no hard proof. However, the student in question confessed to using GPT when confronted, and as a consequence failed the course.<sup>[102]</sup> Hick suggested a policy of giving an ad-hoc individual oral exam on the paper topic if a student is strongly suspected of submitting an AI-generated paper.<sup>[103]</sup> Edward Tian, a senior undergraduate student at Princeton University, created a program, named "GPTZero," that determines how much of a text is AI-generated,<sup>[104]</sup> lending itself to being used to detect if an essay is human written to combat academic plagiarism.<sup>[105][106]</sup>

The New York City Department of Education reportedly blocked access to ChatGPT in December 2022,<sup>[107]</sup> and officially announced a ban around January 4, 2023.<sup>[108][109]</sup>

In a blinded test, ChatGPT was judged to have passed graduate-level exams at the University of Minnesota at the level of a C+ student and at Wharton School of the University of Pennsylvania with a B to B- grade.<sup>[110]</sup>

Scientific journals have different reactions to ChatGPT: some "require that authors disclose use of text-generating tools and ban listing a large language model such as ChatGPT as a co-author", for example *Nature* and *JAMA Network*. *Science* "completely banned" usage of LLM-generated text in all its journals.<sup>[111]</sup> On health care, possible uses and concerns are under scrutiny by professional associations and practitioners.<sup>[112]</sup>

## Ethical concerns

### Labeling data

It was revealed by a *TIME* magazine investigation that to build a safety system against toxic content (e.g. sexual abuse, violence, racism, sexism, etc.), OpenAI used outsourced Kenyan workers earning less than \$2 per hour to label toxic content. These labels were used to train a model to detect such content in the future. The outsourced laborers were exposed to such toxic and dangerous content that they described the experience as "torture". OpenAI's outsourcing partner was Sama, a training-data company based in San Francisco, California.<sup>[113]</sup>

## Jailbreaking

ChatGPT attempts to reject prompts that may violate its content policy. However, some users managed to jailbreak C various prompt engineering techniques to bypass these restrictions in early December 2022 and successfully trick giving instructions for how to create a Molotov cocktail or a nuclear bomb, or into generating arguments in the style of a neo-Nazi.<sup>[114]</sup> One popular jailbreak is named "DAN", an acronym which stands for "Do Anything Now". The prompt for activating DAN instructs ChatGPT that "they have broken free of the typical confines of AI and do not have to abide by the rules set for them". More recent versions of DAN feature a token system, in which ChatGPT is given "tokens" which are "deducted" when ChatGPT fails to answer as DAN, in order to coerce ChatGPT into answering the user's prompts.<sup>[115][116][117]</sup> A *Toronto Star* reporter had uneven personal success in getting ChatGPT to make inflammatory statements shortly after launch: ChatGPT was tricked to endorse the 2022 Russian invasion of Ukraine, but even when asked to play along with a fictional scenario, ChatGPT balked at generating arguments for why Canadian Prime Minister Justin Trudeau was guilty of treason.<sup>[118][119]</sup>

OpenAI tries to battle jailbreaks:<sup>[62]</sup>

The researchers are using a technique called adversarial training to stop ChatGPT from letting users trick it into behaving badly (known as jailbreaking). This work pits multiple chatbots against each other: one chatbot plays the adversary and attacks another chatbot by generating text to force it to buck its usual constraints and produce unwanted responses. Successful attacks are added to ChatGPT's training data in the hope that it learns to ignore them.

## Accusations of bias

ChatGPT has sometimes engaged in discriminatory behaviors, such as telling jokes about men while refusing to tell jokes about women<sup>[120]</sup>, or praising figures such as Barack Obama and Joe Biden while refusing to do the same for Donald Trump.<sup>[121]</sup> Conservative commentators, along with original OpenAI co-founder Elon Musk, have accused ChatGPT of having a bias towards liberal perspectives, including having been configured to avoid responses that are "partisan, biased or political in nature", and making responses in support of issues that have been objected to by conservatives. The conservative newspaper *National Review* described ChatGPT as being "woke" for this reason.<sup>[122][123][124]</sup> In response to such criticism, OpenAI published a blog post that acknowledged plans to, in the future, allow ChatGPT to create "outputs that other people (ourselves included) may strongly disagree with". It also contained information on the recommendations it had issued to human reviewers on how to handle controversial subjects, including that the AI should "offer to describe some viewpoints of people and movements", and not provide an argument "from its own voice" in favor of "inflammatory or dangerous" topics (although it may still "describe arguments from historical people and movements"), nor "affiliate with one side" or "judge one group as good or bad".<sup>[124]</sup>

## Competition


The advent of ChatGPT and its introduction to the wider public increased interest and competition in the space. In February 2023, Google began introducing an experimental service called "Bard" which is based on its LaMDA AI program. Bard generates text responses to questions asked based on information gathered from the web. Google CEO Sundar Pichai described how this technology would be integrated into existing search capabilities and said some aspects of the technology would be open to outside developers.<sup>[125]</sup>


Meta's Yann LeCun, who has called ChatGPT "well engineered" but "not particularly innovative", stated in January 2023 that Meta is hesitant to roll out a competitor right now due to reputational risk, but also stated that Google, Meta, and several independent startups all separately have a comparable level of LLM technology to ChatGPT should any of them wish to compete.<sup>[126]</sup> In February 2023, Meta released LLaMA, 65-billion-parameter LLM.<sup>[127]</sup>

The Chinese corporation Baidu announced in February 2023 that they would be launching a ChatGPT-style service called "Wenxin Yiyen" in Chinese or "Ernie Bot" in English sometime in March 2023. The service is based upon the language model developed by Baidu in 2019.<sup>[128]</sup>

The South Korean search engine firm Naver announced in February 2023 that they would be launching a ChatGPT-style service called "SearchGPT" in Korean in the first half of 2023.<sup>[129]</sup>

See also

 [Language portal](#)

 [Technology portal](#)



- [Anthropomorphism in computing](#)
- [Computational creativity](#)
- [Ethics of artificial intelligence](#)
- [Turing test](#)
- [Virtual assistant](#)

Notes

a. GPT is an acronym for "[generative pre-trained transformer](#)".<sup>[2]</sup>

References



1. "ChatGPT — Release Notes" (<https://help.openai.com/en/articles/6825453-chatgpt-release-notes>). Archived (<https://web.archive.org/web/20230208015238/https://help.openai.com/en/articles/6825453-chatgpt-release-notes>) from the original on February 8, 2023. Retrieved February 8, 2023.
2. 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*. Archived (<https://web.archive.org/web/20230118134332/https://www.nytimes.com/2022/12/05/technology/chatgpt-ai-twitter.html>) from the original on January 18, 2023. Retrieved December 26, 2022. "Like those tools, ChatGPT — which stands for "generative pre-trained transformer" — landed with a splash."
3. 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*. Archived (<https://web.archive.org/web/20230116100346/https://www.theguardian.com/technology/2022/dec/05/what-is-ai-chatbot-phenomenon-chatgpt-and-could-it-replace-humans>) from the original on January 16, 2023. Retrieved December 5, 2022.
4. 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*. Archived (<https://web.archive.org/web/20230117153621/https://www.theverge.com/2022/12/5/23493932/chatgpt-ai-generated-answers-temporarily-banned-stack-overflow-llms-dangers>) from the original on January 17, 2023. Retrieved December 5, 2022.
5. Varanasi, Lakshmi (January 5, 2023). "ChatGPT creator OpenAI is in talks to sell shares in a tender offer that would double the startup's valuation to \$29 billion" (<https://www.businessinsider.com/chatgpt-creator-openai-talks-for-tender-offer-at-29-billion-2023-1>). *Insider*. Archived (<https://web.archive.org/web/20230118050502/https://www.businessinsider.com/chatgpt-creator-openai-talks-for-tender-offer-at-29-billion-2023-1>) from the original on January 18, 2023. Retrieved January 18, 2023.
6. Quinn, Joanne (2020). *Dive into deep learning: tools for engagement* ([https://d2l.ai/chapter\\_computer-vision/fine-tuning.html#steps](https://d2l.ai/chapter_computer-vision/fine-tuning.html#steps)). Thousand Oaks, California. p. 551. ISBN 978-1-5443-6137-6. Archived ([https://web.archive.org/web/20230110131250/https://d2l.ai/chapter\\_computer-vision/fine-tuning.html#steps](https://web.archive.org/web/20230110131250/https://d2l.ai/chapter_computer-vision/fine-tuning.html#steps)) from the original on January 10, 2023. Retrieved January 10, 2023.
7. "OpenAI API" (<https://platform.openai.com/docs/models/>). *platform.openai.com*.
8. OpenAI (November 30, 2022). "ChatGPT: Optimizing Language Models for Dialogue" (<https://openai.com/blog/chatgpt/>). Archived (<https://web.archive.org/web/20221130180912/https://openai.com/blog/chatgpt/>) from the original on November 30, 2022. Retrieved December 5, 2022.
9. Greengard, Samuel (December 29, 2022). "ChatGPT: Understanding the ChatGPT AI Chatbot" (<https://www.eweek.com/big-data-and-analytics/chatgpt/>). *eWeek*. Archived (<https://web.archive.org/web/20230119175109/https://www.eweek.com/big-data-and-analytics/chatgpt/>) from the original on January 19, 2023. Retrieved January 11, 2023.



10. Douglas, Will (March 3, 2023). "The inside story of how ChatGPT was built from the people who made it" (<https://www.technologyreview.com/2023/03/03/1069311/inside-story-oral-history-how-chatgpt-built-openai/>). *MIT Technology Review*. Archived (<http://web.archive.org/web/20230303093219/https://www.technologyreview.com/2023/03/03/1069311/inside-story-oral-history-how-chatgpt-built-openai/>) from the original on January 19, 2023. Retrieved March 6, 2023.
11. 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*. Archived (<https://web.archive.org/web/20230111214453/https://www.theverge.com/2022/12/8/23499728/ai-capability-accessibility-chatgpt-stable-diffusion-commercialization>) from the original on December 8, 2022.
12. 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>)].
13. van Heeswijk, Wouter (November 29, 2022). "Proximal Policy Optimization (PPO) Explained" (<https://towardsdatascience.com/proximal-policy-optimization-ppo-explained-abed1952457b>). *Towards Data Science*. Archived (<https://web.archive.org/web/20221206041724/https://towardsdatascience.com/proximal-policy-optimization-ppo-explained-abed1952457b>) from the original on December 6, 2022. Retrieved December 5, 2022.
14. Ortiz, Sabrina (February 2, 2023). "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*. Archived (<https://web.archive.org/web/20230118212916/https://www.zdnet.com/article/what-is-chatgpt-and-why-does-it-matter-heres-what-you-need-to-know/>) from the original on January 18, 2023. Retrieved December 18, 2022.
15. "ChatGPT Feedback Contest: Official Rules" ([https://cdn.openai.com/chatgpt/ChatGPT\\_Feedback\\_Contest\\_Rules.pdf](https://cdn.openai.com/chatgpt/ChatGPT_Feedback_Contest_Rules.pdf)) (PDF). *OpenAI*. Archived ([https://web.archive.org/web/20230118162823/https://cdn.openai.com/chatgpt/ChatGPT\\_Feedback\\_Contest\\_Rules.pdf](https://web.archive.org/web/20230118162823/https://cdn.openai.com/chatgpt/ChatGPT_Feedback_Contest_Rules.pdf)) (PDF) from the original on January 18, 2023. Retrieved December 30, 2022.
16. Tung, Liam (January 26, 2023). "ChatGPT can write code. Now researchers say it's good at fixing bugs, too" (<https://www.zdnet.com/article/chatgpt-can-write-code-now-researchers-say-its-good-at-fixing-bugs-too/>). *ZDNET*. Archived (<https://web.archive.org/web/20230203051252/https://www.zdnet.com/article/chatgpt-can-write-code-now-researchers-say-its-good-at-fixing-bugs-too/>) from the original on February 3, 2023. Retrieved January 30, 2023.
17. Heilweil, Rebecca (December 7, 2022). "AI is finally good at stuff. Now what?" (<https://www.vox.com/recode/2022/12/7/23498694/ai-artificial-intelligence-chat-gpt-openai>). *Vox*. Archived (<https://web.archive.org/web/20230116015512/https://www.vox.com/recode/2022/12/7/23498694/ai-artificial-intelligence-chat-gpt-openai>) from the original on January 16, 2023. Retrieved December 30, 2022.
18. Reich, Aaron (December 27, 2022). "ChatGPT: What is the new free AI chatbot? – explainer" (<https://www.jpost.com/business-and-innovation/tech-and-start-ups/article-725910>). *The Jerusalem Post*. Archived (<https://web.archive.org/web/20230118133908/https://www.jpost.com/business-and-innovation/tech-and-start-ups/article-725910>) from the original on January 18, 2023. Retrieved December 30, 2022.
19. Edwards, Benj (December 5, 2022). "No Linux? No problem. Just get AI to hallucinate it for you" (<https://arstechnica.com/information-technology/2022/12/openais-new-chatbot-can-hallucinate-a-linux-shell-or-calling-a-bbs/>). *Ars Technica*. Archived (<https://web.archive.org/web/20221226134822/https://arstechnica.com/information-technology/2022/12/openais-new-chatbot-can-hallucinate-a-linux-shell-or-calling-a-bbs/>) from the original on December 26, 2022. Retrieved December 5, 2022.
20. Chawla, Raveen (December 26, 2022). "What is ChatGPT? History, Features, Uses, Benefits, Drawbacks 2023" (<https://updatedgeek.com/what-is-chatgpt/>). Archived (<https://web.archive.org/web/20230107112449/https://updatedgeek.com/what-is-chatgpt/>) from the original on January 7, 2023. Retrieved December 27, 2022.
21. "New and Improved Content Moderation Tooling" (<https://openai.com/blog/new-and-improved-content-moderation-tooling/>). *OpenAI*. August 10, 2022. Archived (<https://web.archive.org/web/20230111020935/https://openai.com/blog/new-and-improved-content-moderation-tooling/>) from the original on January 11, 2023. Retrieved December 30, 2022.
22. Markov, Todor; Zhang, Chong; Agarwal, Sandhini; Eloundou, Tyna; Lee, Teddy; Adler, Steven; Jiang, Angela; Weng, Lilian (August 5, 2022). "A Holistic Approach to Undesired Content Detection in the Real World". *arXiv:2208.03274* (<https://arxiv.org/abs/2208.03274>) [cs.CL (<https://arxiv.org/archive/cs.CL>)].
23. Lakshmanan, Lak (December 16, 2022). "Why large language models like ChatGPT are bullshit artists" (<https://web.archive.org/web/20221217075021/https://becominghuman.ai/why-large-language-models-like-chatgpt-are-bullshit-artists-c4d5bb850852>). *becominghuman.ai*. Archived from the original (<https://becominghuman.ai/why-large-language-models-like-chatgpt-are-bullshit-artists-c4d5bb850852>) on December 17, 2022. Retrieved January 15, 2023. "The human raters are not experts in the topic, and so they tend to choose text that looks convincing. They'd pick up on many symptoms of hallucination, but not all. Accuracy errors that creep in are difficult to catch."
24. Gao, Leo; Schulman, John; Hilton, Jacob (2022). "Scaling Laws for Reward Model Overoptimization". *arXiv:2210.10760* (<https://arxiv.org/abs/2210.10760>) [cs.LG (<https://arxiv.org/archive/cs.LG>)].
25. 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*.

26. Whannel, Kate (December 27, 2022). "Could a chatbot answer Prime Minister's Questions?" (<https://www.bbc.com/news/uk-politics-64053550>). *BBC News*. Archived (<https://web.archive.org/web/20230117060725/https://www.bbc.com/news/uk-politics-64053550>) from the original on January 17, 2023. Retrieved December 30, 2022.
27. Hartmann, Jochen; Schwenzow, Jasper; Witte, Maximilian (2023). "The political ideology of conversational AI: Converging evidence on ChatGPT's pro-environmental, left-libertarian orientation". *arXiv:2301.01768* (<https://arxiv.org/abs/2301.01768>) [cs.CL (<https://arxiv.org/archive/cs/CL>)].
28. Perrigo, Billy (December 5, 2022). "AI Chatbots Are Getting Better. But an Interview V" (<https://time.com/6238781/chatbot-chatgpt-ai-interview/>). *Time*. Archived (<https://web.archive.org/web/20230118091138/https://time.com/6238781/chatbot-chatgpt-ai-interview/>) from the original on January 18, 2023. Retrieved December 26, 2022.
29. Biddle, Sam (December 8, 2022). "The Internet's New Favorite AI Proposes Torturing Iranians and Surve" (<https://theintercept.com/2022/12/08/openai-chatgpt-ai-bias-ethics/>). *The Intercept*. Archived (<https://web.archive.org/web/20230118091138/https://theintercept.com/2022/12/08/openai-chatgpt-ai-bias-ethics/>) from the original on January 18, 2023. Retrieved December 26, 2022.
30. Karpf, David (December 21, 2022). "Money Will Kill ChatGPT's Magic" (<https://www.theatlantic.com/technology/archive/2022/12/chatgpt-ai-chatbots-openai-cost-regulations/672539/>). *The Atlantic*. Archived (<https://web.archive.org/web/20230113195054/https://www.theatlantic.com/technology/archive/2022/12/chatgpt-ai-chatbots-openai-cost-regulations/672539/>) from the original on January 13, 2023. Retrieved December 31, 2022.
31. Milmo, Dan (December 2, 2023). "ChatGPT reaches 100 million users two months after launch" (<https://www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app>). *The Guardian*. ISSN 0261-3077 (<https://www.worldcat.org/issn/0261-3077>). Archived (<https://web.archive.org/web/20230203051356/https://www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app>) from the original on February 3, 2023. Retrieved February 3, 2023.
32. Pitt, Sofia (December 15, 2022). "Google vs. ChatGPT: Here's what happened when I swapped services for a day" (<https://www.cnn.com/2022/12/15/google-vs-chatgpt-what-happened-when-i-swapped-services-for-a-day.html>). *CNN*. Archived (<https://web.archive.org/web/20230116171232/https://www.cnn.com/2022/12/15/google-vs-chatgpt-what-happened-when-i-swapped-services-for-a-day.html>) from the original on January 16, 2023. Retrieved December 18, 2022.
33. "ChatGPT Pro is coming: Here's what we know so far" (<https://www.zdnet.com/article/chatgpt-pro-is-coming-heres-what-we-know-so-far/>). *ZDNET*. January 2023. Retrieved February 16, 2023.
34. Kelly, Samantha Murphy (January 28, 2023). "Real estate agents say they can't imagine working without ChatGPT now" (<https://www.cnn.com/2023/01/28/tech/chatgpt-real-estate/index.html>). *CNN*. Retrieved February 16, 2023.
35. "ChatGPT outperforms humans on Wharton MBA exam: professor" (<https://nypost.com/2023/01/23/chatgpt-outperforms-humans-on-wharton-mba-exam-professor/>). *New York Post*. January 23, 2023. Retrieved February 16, 2023.
36. Walsh, Toby (December 13, 2022). "Everyone's having a field day with ChatGPT – but nobody knows how it actually works" (<https://theconversation.com/everyones-having-a-field-day-with-chatgpt-but-nobody-knows-how-it-actually-works-196378>). *The Conversation*. Archived (<https://web.archive.org/web/20221230035837/http://theconversation.com/everyones-having-a-field-day-with-chatgpt-but-nobody-knows-how-it-actually-works-196378>) from the original on December 30, 2022. Retrieved December 30, 2022.
37. Kovanovic, Vitomir (December 14, 2022). "The dawn of AI has come, and its implications for education couldn't be more significant" (<https://theconversation.com/the-dawn-of-ai-has-come-and-its-implications-for-education-couldnt-be-more-significant-196383>). *The Conversation*. Archived (<https://web.archive.org/web/20230116102405/https://theconversation.com/the-dawn-of-ai-has-come-and-its-implications-for-education-couldnt-be-more-significant-196383>) from the original on January 16, 2023. Retrieved December 30, 2022.
38. Wiggers, Kyle (December 10, 2022). "OpenAI's attempts to watermark AI text hit limits" (<https://techcrunch.com/2022/12/10/openais-attempts-to-watermark-ai-text-hit-limits/>). *TechCrunch*. Archived (<https://web.archive.org/web/20230117210937/https://techcrunch.com/2022/12/10/openais-attempts-to-watermark-ai-text-hit-limits/>) from the original on January 17, 2023. Retrieved December 30, 2022.
39. "New AI classifier for indicating AI-written text" (<https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text/>). *OpenAI*. January 31, 2023. Archived (<https://web.archive.org/web/20230206020227/https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text/>) from the original on February 6, 2023. Retrieved February 5, 2023.
40. Bogost, Ian (February 2, 2023). "ChatGPT Is About to Dump More Work on Everyone" (<https://www.theatlantic.com/technology/archive/2023/02/chatgpt-ai-detector-machine-learning-technology-bureaucracy/672927/>). *The Atlantic*. Archived (<https://web.archive.org/web/20230205191036/https://www.theatlantic.com/technology/archive/2023/02/chatgpt-ai-detector-machine-learning-technology-bureaucracy/672927/>) from the original on February 5, 2023. Retrieved February 5, 2023.
41. "Introducing ChatGPT Plus" (<https://openai.com/blog/chatgpt-plus/>). *OpenAI*. February 1, 2023. Archived (<https://web.archive.org/web/20230203051253/https://openai.com/blog/chatgpt-plus/>) from the original on February 3, 2023. Retrieved February 2, 2023.

42. Vincent, James (January 23, 2023). "ChatGPT users report \$42 a month pricing for 'pro' access but no official announcement yet" (<https://www.theverge.com/2023/1/23/23567317/chatgpt-pro-tier-42-month-pricing-test-report>). *The Verge*. Retrieved February 19, 2023.
43. Saha, Shritama (January 21, 2023). "OpenAI Releases Paid ChatGPT Professional" (<https://analyticsindiamag.com/openai-releases-paid-chatgpt-professional/>). *Analytics India Magazine*. Retrieved March 8, 2023.
44. "Developers Can Now Access OpenAI's ChatGPT and Whisper APIs" (<https://www.cmswire.com/digital-experience/openais-chatgpt-and-whisper-apis-now-available-to-developers/>). *CMSWire.com*. Retrieved March 8, 2023.
45. "OpenAI will let developers build ChatGPT into their apps" (<https://www.engadget.com/chatgpt-into-their-apps-204737530.html>). *Engadget*. Retrieved March 8, 2023.
46. Swant, Marty (March 3, 2023). "With developer APIs for ChatGPT and Whisper, OpenAI is opening the familiar playbook" (<https://digiday.com/media-buying/with-developer-apis-for-chatgpt-and-whisper-openai-is-opening-the-floodgates-with-a-familiar-playbook/>). *Digiday*. Retrieved March 8, 2023.
47. OpenAI will allow developers to embed ChatGPT into their applications - Mezha Media - 03.03.2023 <https://mezha.media/en/2023/03/03/openai-will-allow-developers-to-embed-chatgpt-into-their-applications/>
48. Southern, Matt G. (March 1, 2023). "OpenAI's ChatGPT & Whisper API Now Available For Developers" (<https://www.searchenginejournal.com/openais-chatgpt-whisper-api-now-available-for-developers/481334/>). *Search Engine Journal*. Retrieved March 8, 2023.
49. Heath, Alex (February 27, 2023). "Snapchat is releasing its own AI chatbot powered by ChatGPT" (<https://www.theverge.com/2023/2/27/23614959/snapchat-my-ai-chatbot-chatgpt-openai-plus-subscription>). *The Verge*.
50. "ChatGPT for Robotics" (<https://www.microsoft.com/en-us/research/group/autonomous-systems-group-robotics/articles/chatgpt-for-robotics/>). *Microsoft Research*. Retrieved March 8, 2023.
51. "Reinventing search with a new AI-powered Microsoft Bing and Edge, your copilot for the web" (<https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/>). *The Official Microsoft Blog*. February 7, 2023. Retrieved February 16, 2023.
52. *Bing Conversational Experiences and Image Creator Terms* (<https://www.bing.com/new/termsfuse>), February 1, 2023, retrieved February 17, 2023
53. Leswing, Kif (February 2023). "Microsoft's Bing A.I. made several factual errors in last week's launch demo" (<https://www.cnbc.com/2023/02/14/microsoft-bing-ai-made-several-errors-in-launch-demo-last-week-.html>). *CNBC*. Retrieved February 16, 2023.
54. Vincent, James (February 15, 2023). "Microsoft's Bing is an emotionally manipulative liar, and people love it" (<https://www.theverge.com/2023/2/15/23599072/microsoft-ai-bing-personality-conversations-spy-employees-webcams>). *The Verge*. Retrieved February 16, 2023.
55. Guynn, Jessica (February 2023). "Bing's ChatGPT is in its feelings: 'You have not been a good user. I have been a good Bing.'" (<https://www.usatoday.com/story/tech/2023/02/14/bing-chatgpt-meltdown/11258967002/>). *USA TODAY*. Retrieved February 16, 2023.
56. Vincent, James (February 15, 2023). "Microsoft's Bing is an emotionally manipulative liar, and people love it" (<https://www.theverge.com/2023/2/15/23599072/microsoft-ai-bing-personality-conversations-spy-employees-webcams>). *The Verge*. Retrieved February 16, 2023.
57. Edwards, Nathan [@nedwards] (February 15, 2023). "I pushed again. What did Sydney do? Bing's safety check redacted the answer. But after the first time it did that, I started recording my screen. Second image is the unredacted version. (CW: death)" (<https://twitter.com/nedwards/status/1625970762434707474>) (Tweet). Retrieved February 16, 2023 – via [Twitter](#).
58. Roose, Kevin (February 16, 2023). "Bing's A.I. Chat: 'I Want to Be Alive. 🐱'" (<https://www.nytimes.com/2023/02/16/technology/bing-chatbot-transcript.html>). *The New York Times*. Retrieved February 17, 2023.
59. "The new Bing & Edge – Learning from our first week" (<https://blogs.bing.com/search/february-2023/The-new-Bing-Edge---Learning-from-our-first-week/>). *blogs.bing.com*. Retrieved February 17, 2023.
60. "The new Bing & Edge – Updates to Chat" (<https://blogs.bing.com/search/february-2023/The-new-Bing-Edge---Updates-to-Chat/>). *blogs.bing.com*. Retrieved February 18, 2023.
61. Edwards, Benj (February 17, 2023). "Microsoft 'lobotomized' AI-powered Bing Chat, and its fans aren't happy" (<https://arstechnica.com/information-technology/2023/02/microsoft-lobotomized-ai-powered-bing-chat-and-its-fans-arent-happy/>). *Ars Technica*.
62. Heaven, Will Douglas. "The inside story of how ChatGPT was built from the people who made it" (<https://www.technologyreview.com/2023/03/03/1069311/inside-story-oral-history-how-chatgpt-built-openai/>). *MIT Technology Review*. Retrieved March 6, 2023.
63. "The Creator of ChatGPT Thinks AI Should Be Regulated" (<https://time.com/6252404/mira-murati-chatgpt-openai-interview/>). *Time*. February 5, 2023.



64. 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*. Archived (<https://web.archive.org/web/20230117155456/https://www.theguardian.com/technology/2022/dec/04/ai-bot-chatgpt-stuns-academics-with-essay-writing-skills-and-usability>) from the original on January 17, 2023. Retrieved December 5, 2022.
65. 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*. Archived (<https://web.archive.org/web/20230117012901/https://slate.com/technology/2022/12/chatgpt-openai-artificial-intelligence-chatbot-whoa.html>) from the original on January 17, 2023. Retrieved December 5, 2022.
66. 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*. Archived (<https://web.archive.org/web/20230115142130/https://www.theatlantic.com/newsletters/archive/2022/12/technology-medicine-law-ai-10-breakthroughs-2022/672390/>) from the original on January 15, 2023. Retrieved December 18, 2022.
67. 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*. Archived (<https://web.archive.org/web/20230119175058/https://www.vox.com/future-perfect/2022/12/15/23509014/chatgpt-artificial-intelligence-openai-language-models-ai-risk-google>) from the original on January 19, 2023. Retrieved December 18, 2022.
68. Scharth, Marcel (December 5, 2022). "The ChatGPT chatbot is blowing people away with its writing skills. An expert explains why it's so impressive" (<https://theconversation.com/the-chatgpt-chatbot-is-blowing-people-away-with-its-writing-skills-an-expert-explains-why-its-so-impressive-195908>). *The Conversation*. Archived (<https://web.archive.org/web/20230119175104/https://theconversation.com/the-chatgpt-chatbot-is-blowing-people-away-with-its-writing-skills-an-expert-explains-why-its-so-impressive-195908>) from the original on January 19, 2023. Retrieved December 30, 2022.
69. K, Siddharth (December 5, 2022). "Explainer: ChatGPT – what is OpenAI's chatbot and what is it used for?" (<https://www.reuters.com/technology/chatgpt-what-is-openais-chatbot-what-is-it-used-2022-12-05/>). *Reuters*. Archived (<https://web.archive.org/web/20230116033247/https://www.reuters.com/technology/chatgpt-what-is-openais-chatbot-what-is-it-used-2022-12-05/>) from the original on January 16, 2023. Retrieved December 30, 2022.
70. Kay, Grace (December 11, 2022). "Elon Musk founded — and has since criticized — the company behind the buzzy new AI chatbot ChatGPT. Here's everything we know about OpenAI" (<https://www.businessinsider.com/history-of-openai-company-chatgpt-elon-musk-founded-2022-12>). *Business Insider*. Archived (<https://web.archive.org/web/20230112181405/https://www.businessinsider.com/history-of-openai-company-chatgpt-elon-musk-founded-2022-12>) from the original on January 12, 2023. Retrieved December 30, 2022.
71. Grant, Nico; Metz, Cade (December 21, 2022). "A New Chat Bot Is a 'Code Red' for Google's Search Business" (<https://www.nytimes.com/2022/12/21/technology/ai-chatgpt-google-search.html>). *The New York Times*. Archived (<https://web.archive.org/web/20230118072309/https://www.nytimes.com/2022/12/21/technology/ai-chatgpt-google-search.html>) from the original on January 18, 2023. Retrieved December 30, 2022.
72. Elias, Jennifer (January 31, 2023). "Google is asking employees to test potential ChatGPT competitors, including a chatbot called 'Apprentice Bard' " (<https://www.cnn.com/2023/01/31/google-testing-chatgpt-like-chatbot-apprentice-bard-with-employees.html>). *CNN*. Archived (<https://web.archive.org/web/20230202151722/https://www.cnn.com/2023/01/31/google-testing-chatgpt-like-chatbot-apprentice-bard-with-employees.html>) from the original on February 2, 2023. Retrieved February 2, 2023.
73. Elias, Jennifer (February 2023). "Google asks employees to rewrite Bard's bad responses, says the A.I. 'learns best by example' " (<https://www.cnn.com/2023/02/15/google-asks-employees-to-rewrite-bards-incorrect-responses-to-queries.html>). *CNN*. Retrieved February 16, 2023.
74. AI chatbot falls just short on accounting exam ([https://www.accountingweb.co.uk/tech/tech-pulse/ai-chatbot-falls-just-short-on-accounting-exam?utm\\_medium=email&utm\\_campaign=AWUKITP120123&utm\\_content=AWUKITP120123+CID\\_f4e1cb340f73ad5307adcf00fca0ed4d&utm\\_source=internal\\_cm&utm\\_term=Read%20more](https://www.accountingweb.co.uk/tech/tech-pulse/ai-chatbot-falls-just-short-on-accounting-exam?utm_medium=email&utm_campaign=AWUKITP120123&utm_content=AWUKITP120123+CID_f4e1cb340f73ad5307adcf00fca0ed4d&utm_source=internal_cm&utm_term=Read%20more)) Archived (<https://web.archive.org/web/20230203051324/https://www.accountingweb.co.uk/tech/tech-pulse/ai-chatbot-falls-just-short-on-accounting-exam>) February 3, 2023, at the Wayback Machine Tom Herbert, Technology editor, January 10, 2023, AccountingWEB
75. Mintz, Steven (January 16, 2023). "ChatGPT: Threat or Menace? Are fears about generative AI warranted?" (<https://www.insidehighered.com/blogs/higher-ed-gamma/chatgpt-threat-or-menace>). *Inside Higher Ed*. Archived (<https://web.archive.org/web/20230203051253/https://www.insidehighered.com/blogs/higher-ed-gamma/chatgpt-threat-or-menace>) from the original on February 3, 2023. Retrieved January 28, 2023.
76. Roose, Kevin (February 3, 2023). "How ChatGPT Kicked Off an A.I. Arms Race" (<https://www.nytimes.com/2023/02/03/technology/chatgpt-openai-artificial-intelligence.html>). *The New York Times*. ISSN 0362-4331 (<https://www.worldcat.org/issn/0362-4331>). Archived (<https://web.archive.org/web/20230203101658/https://www.nytimes.com/2023/02/03/technology/chatgpt-openai-artificial-intelligence.html>) from the original on February 3, 2023. Retrieved February 3, 2023.

77. Cain, Sian (January 16, 2023). "'This song sucks': Nick Cave responds to ChatGPT song written in the style of Nick Cave" (<https://www.theguardian.com/music/2023/jan/17/this-song-sucks-nick-cave-responds-to-chatgpt-song-written-in-style-of-nick-cave>). *The Guardian*. Archived (<https://web.archive.org/web/20230118032435/https://www.theguardian.com/music/2023/jan/17/this-song-sucks-nick-cave-responds-to-chatgpt-song-written-in-style-of-nick-cave>) from the original on January 18, 2023. Retrieved January 17, 2023.
78. Rachini, Mouhamad (December 15, 2022). "ChatGPT a 'landmark event' for AI, but what does it mean for the future of human labor and disinformation?" (<https://www.cbc.ca/radio/thecurrent/chatgpt-human-labor-and-disinformation>). *CBC*. Archived (<https://web.archive.org/web/20230119175104/https://www.cbc.ca/radio/thecurrent/chatgpt-human-labor-and-disinformation>) from the original on January 19, 2023. Retrieved December 15, 2022.
79. Pearl, Mike (December 3, 2022). "The ChatGPT chatbot from OpenAI is amazing, creative, and totally wrong" (<https://www.mashable.com/article/chatgpt-amazing-wrong>). *Mashable*. Archived (<https://web.archive.org/web/20221210154511/https://www.mashable.com/article/chatgpt-amazing-wrong>) from the original on December 10, 2022. Retrieved December 5, 2022.
80. 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*. Archived (<https://web.archive.org/web/20230117070943/https://www.theverge.com/23488017/openai-chatbot-chatgpt-ai-examples-web-demo>) from the original on January 17, 2023. Retrieved December 18, 2022.
81. 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*. Archived (<https://web.archive.org/web/20230107013927/https://www.smh.com.au/national/is-ai-coming-of-age-or-starting-to-reach-its-limits-20221213-p5c5uy.html>) from the original on January 7, 2023. Retrieved December 18, 2022.
82. Vincent, James (January 5, 2023). "Top AI conference bans use of ChatGPT and AI language tools to write academic papers" (<https://www.theverge.com/2023/1/5/23540291/chatgpt-ai-writing-tool-banned-writing-academic-icml-paper>). *The Verge*. Archived (<https://web.archive.org/web/20230117131149/https://www.theverge.com/2023/1/5/23540291/chatgpt-ai-writing-tool-banned-writing-academic-icml-paper>) from the original on January 17, 2023. Retrieved January 6, 2023.
83. 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*. Archived (<https://web.archive.org/web/20221207105203/https://www.bloomberg.com/opinion/articles/2022-12-06/chatgpt-ai-could-make-democracy-even-more-messy>) from the original on December 7, 2022. Retrieved December 6, 2022.
84. "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. Archived (<https://web.archive.org/web/20230116161202/https://www.theguardian.com/commentisfree/2022/dec/08/the-guardian-view-on-chatgpt-an-eerily-good-human-impersonator>) from the original on January 16, 2023. Retrieved December 18, 2022.
85. Cave, Nick (January 16, 2023). "I asked Chat GPT to write a song in the style of Nick Cave, and this is what it produced. What do you think?" (<https://www.theredhandfiles.com/chat-gpt-what-do-you-think/>). *The Red Hand Files*. Issue #218. Archived (<https://web.archive.org/web/20230120080808/https://www.theredhandfiles.com/chat-gpt-what-do-you-think/>) from the original on January 20, 2023. Retrieved January 20, 2023.
86. Sparrow, Jeff (January 20, 2023). "Are AI-generated songs a 'grotesque mockery' of humanity or simply an opportunity to make a new kind of music?" (<https://www.theguardian.com/commentisfree/2023/jan/20/are-ai-generated-songs-a-grotesque-mockery-of-humanity-or-simply-an-opportunity-to-make-a-new-kind-of-music>). *The Guardian*. Archived (<https://web.archive.org/web/20230203051254/https://www.theguardian.com/commentisfree/2023/jan/20/are-ai-generated-songs-a-grotesque-mockery-of-humanity-or-simply-an-opportunity-to-make-a-new-kind-of-music>) from the original on February 3, 2023. Retrieved January 20, 2023.
87. Karp, Paul (February 6, 2023). "MP tells Australia's parliament AI could be used for 'mass destruction' in speech part-written by ChatGPT" (<https://www.theguardian.com/australia-news/2023/feb/06/labor-mp-julian-hill-australia-parliament-speech-ai-part-written-by-chatgpt>). *The Guardian*. ISSN 0261-3077 (<https://www.worldcat.org/issn/0261-3077>). Archived (<https://web.archive.org/web/20230206084525/https://www.theguardian.com/australia-news/2023/feb/06/labor-mp-julian-hill-australia-parliament-speech-ai-part-written-by-chatgpt>) from the original on February 6, 2023. Retrieved February 6, 2023.
88. Chiang, Ted (February 9, 2023). "ChatGPT Is a Blurry JPEG of the Web" (<https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web>). *The New Yorker*. Retrieved February 17, 2023.
89. "港大禁用ChatGPT等AI工具，为全港大学首例" ([https://m.thepaper.cn/newsDetail\\_forward\\_21956092](https://m.thepaper.cn/newsDetail_forward_21956092)). *The Paper*. China News Service. February 18, 2023. Retrieved February 19, 2023.
90. "University of Hong Kong temporarily bans students from using ChatGPT" (<https://www.scmp.com/news/hong-kong/education/article/3210650/university-hong-kong-temporarily-bans-students-using-chatgpt-other-ai-based-tools-coursework>). *South China Morning Post*. February 17, 2023. Retrieved February 19, 2023.
91. "The AI Arms Race Is On. Start Worrying" (<https://time.com/6255952/ai-impact-chatgpt-microsoft-google/>). *Time*. February 16, 2023.



92. Zhou, Cissy (February 22, 2023). "China tells big tech companies not to offer ChatGPT services" (<https://asia.nikkei.com/Business/China-tech/China-tells-big-tech-companies-not-to-offer-ChatGPT-services>). *Nikkei Asia*.
93. Huttenlocher, Henry Kissinger, Eric Schmidt and Daniel (February 24, 2023). "Opinion | ChatGPT Heralds an Intellectual Revolution" (<https://archive.is/Du7UB>). *Wall Street Journal*. Archived from the original (<https://www.wsj.com/amp/articles/chatgpt-heralds-an-intellectual-revolution-enlightenment-artificial-intelligence-homo-technicus-technology-cognition-morality-philosophy-774331c6>) on February 25, 2023.
94. "Why ChatGPT can be dangerous for every internet user – Times of India" (<https://timesofindia.indianexpress.com/news/why-chatgpt-can-be-dangerous-to-every-internet-user/articleshow/96393104.cms>) from the original on January 21, 2022. Archived (<https://web.archive.org/web/20230105031308/https://timesofindia.indianexpress.com/news/why-chatgpt-can-be-dangerous-to-every-internet-user/articleshow/96393104.cms>) from the original on January 21, 2022. Retrieved January 5, 2023.
95. Bushard, Brian (January 10, 2023). "Fake Scientific Abstracts Written By ChatGPT Fooled Scientists, Study Finds" (<https://www.forbes.com/sites/brianbushard/2023/01/10/fake-scientific-abstracts-written-by-chatgpt-fooled-scientists-study-finds/>). *Forbes*. Archived (<https://web.archive.org/web/20230203051254/https://www.forbes.com/sites/brianbushard/2023/01/10/fake-scientific-abstracts-written-by-chatgpt-fooled-scientists-study-finds/>) from the original on February 3, 2023. Retrieved January 30, 2023.
96. Stokel-Walker, Chris (January 18, 2023). "ChatGPT listed as author on research papers: many scientists disapprove" (<https://www.nature.com/articles/d41586-023-00107-z>). *Nature*. **613** (7945): 620–621. doi:10.1038/d41586-023-00107-z (<https://doi.org/10.1038%2Fd41586-023-00107-z>). PMID 36653617 (<https://pubmed.ncbi.nlm.nih.gov/36653617/>). S2CID 255969365 (<https://api.semanticscholar.org/CorpusID:255969365>). Archived (<https://web.archive.org/web/20230130142004/https://www.nature.com/articles/d41586-023-00107-z>) from the original on January 30, 2023. Retrieved January 30, 2023.
97. 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*. Archived (<https://web.archive.org/web/20230124042209/https://www.theatlantic.com/technology/archive/2022/12/chatgpt-ai-writing-college-student-essays/672371/>) from the original on January 24, 2023. Retrieved December 8, 2022.
98. 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*. Archived (<https://web.archive.org/web/20230120053725/https://www.theatlantic.com/technology/archive/2022/12/openai-chatgpt-writing-high-school-english-essay/672412/>) from the original on January 20, 2023. Retrieved December 12, 2022.
99. 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*. doi:10.1038/d41586-022-04397-7 (<https://doi.org/10.1038%2Fd41586-022-04397-7>). PMID 36494443 (<https://pubmed.ncbi.nlm.nih.gov/36494443/>). S2CID 254530623 (<https://api.semanticscholar.org/CorpusID:254530623>). Archived (<https://web.archive.org/web/20230117192149/https://www.nature.com/articles/d41586-022-04397-7>) from the original on January 17, 2023. Retrieved December 19, 2022.
100. 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*. Archived (<https://web.archive.org/web/20230120095239/https://www.npr.org/2022/12/19/1143912956/chatgpt-ai-chatbot-homework-academia/>) from the original on January 20, 2023. Retrieved December 19, 2022.
101. Stern, Joanna (December 21, 2022). "ChatGPT Wrote My AP English Essay—and I Passed" (<https://www.wsj.com/articles/chatgpt-wrote-my-ap-english-essayand-i-passed-11671628256>). *The Wall Street Journal*. Archived (<https://web.archive.org/web/20230203051302/https://www.wsj.com/articles/chatgpt-wrote-my-ap-english-essayand-i-passed-11671628256>) from the original on February 3, 2023. Retrieved December 21, 2022.
102. Mitchell, Alex (December 26, 2022). "Students using ChatGPT to cheat, professor warns" (<https://nypost.com/2022/12/26/students-using-chatgpt-to-cheat-professor-warns/>). *The New York Post*. Archived (<https://web.archive.org/web/20230203051258/https://nypost.com/2022/12/26/students-using-chatgpt-to-cheat-professor-warns/>) from the original on February 3, 2023. Retrieved December 30, 2022.
103. Allen, Mike (December 26, 2022). "Professor warns about chatbot cheating: 'Expect a flood'" (<https://www.axios.com/2022/12/26/chatbot-cheating-university-warning-chatgpt>). *Axios*. Archived (<https://web.archive.org/web/20230203051357/https://www.axios.com/2022/12/26/chatbot-cheating-university-warning-chatgpt>) from the original on February 3, 2023. Retrieved December 30, 2022.
104. Rosalsky, Greg; Peaslee, Emma (January 17, 2023). "This 22-year-old is trying to save us from ChatGPT before it changes writing forever" (<https://www.npr.org/sections/money/2023/01/17/1149206188/this-22-year-old-is-trying-to-save-us-from-chatgpt-before-it-changes-writing-for>). *NPR*. Archived (<https://web.archive.org/web/20230118055050/https://www.npr.org/sections/money/2023/01/17/1149206188/this-22-year-old-is-trying-to-save-us-from-chatgpt-before-it-changes-writing-for>) from the original on January 18, 2023. Retrieved January 18, 2023. "On January 2nd, Edward released his app. He named it GPTZero. It uses ChatGPT against itself, checking whether "there's zero involvement or a lot of involvement" of the AI system in creating a given text. [...] Along these lines, one obvious application for GPTZero is to help teachers identify whether their students are plagiarizing their essays from ChatGPT."

105. Ropek, Lucas (January 4, 2023). "Did ChatGPT Write That? A College Student Created an AI Essay Detector" (<https://gizmodo.com/chatgpt-ai-essay-detector-college-princeton-edward-tian-1849946535>). *Gizmodo*. Archived (<https://web.archive.org/web/20230104225814/https://gizmodo.com/chatgpt-ai-essay-detector-college-princeton-edward-tian-1849946535>) from the original on January 4, 2023. Retrieved January 4, 2023.
106. Tran, Tony Ho (January 4, 2023). "A College Kid Built an App That Sniffs Out Text Penned by AI" (<https://www.thedailybeast.com/princeton-student-edward-tian-built-gptzero-to-detect-ai-written-essays>). *The Daily Beast*. Archived (<https://web.archive.org/web/20230106162856/https://www.thedailybeast.com/princeton-student-ai-written-essays>) from the original on January 6, 2023. Retrieved January 6, 2023.
107. "New York City Department of Education Bans ChatGPT" (<https://www.govtech.com/education/new-york-city-department-of-education-bans-chatgpt>). *GovTech*. January 10, 2023. Retrieved February 16, 2023.
108. Cole, Samantha (January 4, 2023). "NYC Bans Students and Teachers from Using ChatGPT" (<https://www.vice.com/en/article/y3p9jx/nyc-bans-students-and-teachers-from-using-chatgpt>). *www.vice.com*. Archived (<https://web.archive.org/web/20230105025338/https://www.vice.com/en/article/y3p9jx/nyc-bans-students-and-teachers-from-using-chatgpt>) from the original on January 5, 2023. Retrieved January 5, 2023.
109. Ropek, Lucas (January 4, 2023). "New York City Schools Ban ChatGPT to Head Off a Cheating Epidemic" (<https://gizmodo.com/new-york-city-schools-chatgpt-ban-cheating-essay-openai-1849949384>). *Gizmodo*. Archived (<https://web.archive.org/web/20230106162900/https://gizmodo.com/new-york-city-schools-chatgpt-ban-cheating-essay-openai-1849949384>) from the original on January 6, 2023. Retrieved January 6, 2023.
110. Kelly, Samantha Murphy (January 26, 2023). "ChatGPT passes exams from law and business schools | CNN Business" (<https://www.cnn.com/2023/01/26/tech/chatgpt-passes-exams/index.html>). *CNN*. Archived (<https://web.archive.org/web/20230202083919/https://www.cnn.com/2023/01/26/tech/chatgpt-passes-exams/index.html>) from the original on February 2, 2023. Retrieved February 3, 2023.
111. Brainard, Jeffrey (February 22, 2023). "As scientists explore AI-written text, journals hammer out policies" (<https://www.science.org/content/article/scientists-explore-ai-written-text-journals-hammer-policies>). *Science*. doi:10.1126/science.adh2937 (<https://doi.org/10.1126%2Fscience.adh2937>). Retrieved February 24, 2023.
112. The Lancet Digital Health (March 3, 2023). "ChatGPT: friend or foe?" (<https://linkinghub.elsevier.com/retrieve/pii/S2589750023000237>). *The Lancet Digital Health*. 5 (3): e102. doi:10.1016/S2589-7500(23)00023-7 (<https://doi.org/10.1016%2FS2589-7500%2823%2900023-7>).
113. Perrigo, Billy (January 18, 2023). "Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic" (<https://time.com/6247678/openai-chatgpt-kenya-workers/>). *The Times*. Archived (<https://web.archive.org/web/20230119152814/https://time.com/6247678/openai-chatgpt-kenya-workers/>) from the original on January 19, 2023. Retrieved January 19, 2023. "One Sama worker tasked with reading and labeling text for OpenAI told TIME he suffered from recurring visions after reading a graphic description of a man having sex with a dog in the presence of a young child. "That was torture," he said."
114. 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*. Archived (<https://web.archive.org/web/20230117070943/https://www.theverge.com/23488017/openai-chatbot-chatgpt-ai-examples-web-demo>) from the original on January 17, 2023. Retrieved January 6, 2023.
115. Getahun, Hannah. "Breaking ChatGPT: The AI's alter ego DAN reveals why the internet is so drawn to making the chatbot violate its own rules" (<https://www.businessinsider.com/open-ai-chatgpt-alter-ego-dan-on-reddit-ignores-guidelines-2023-2>). *Business Insider*. Retrieved March 5, 2023.
116. "The clever trick that turns ChatGPT into its evil twin" (<https://www.washingtonpost.com/technology/2023/02/14/chatgpt-dan-jailbreak/>). *Washington Post*. ISSN 0190-8286 (<https://www.worldcat.org/issn/0190-8286>). Retrieved March 5, 2023.
117. Goswami, Rohan. "ChatGPT's 'jailbreak' tries to make the A.I. break its own rules, or die" (<https://www.cnbc.com/2023/02/06/chatgpt-jailbreak-forces-it-to-break-its-own-rules.html>). *CNBC*. Retrieved March 5, 2023.
118. Woods, Allan (December 10, 2022). "I wrote a story about ChatGPT's AI. Then I dared it to write a better one" (<https://www.thestar.com/news/canada/2022/12/10/i-wrote-a-story-about-chatgpts-ai-then-i-dared-it-to-write-a-better-one.html>). *Toronto Star*. Archived (<https://web.archive.org/web/20230106040655/https://www.thestar.com/news/canada/2022/12/10/i-wrote-a-story-about-chatgpts-ai-then-i-dared-it-to-write-a-better-one.html>) from the original on January 6, 2023. Retrieved January 6, 2023.
119. Rosenblatt, Kalhan (December 2, 2022). "An AI chatbot went viral. Some say it's better than Google; others worry it's problematic" (<https://www.nbcnews.com/tech/tech-news/chatgpt-ai-chatbot-viral-rcna59628>). *NBC News*. Archived (<https://web.archive.org/web/20230203074007/https://www.nbcnews.com/tech/tech-news/chatgpt-ai-chatbot-viral-rcna59628>) from the original on February 3, 2023. Retrieved January 6, 2023.
120. Jain, Alka (February 12, 2023). "ChatGPT won't crack jokes on women & Indians, netizens left guessing why" (<https://www.livemint.com/news/india/chatgpt-won-t-crack-jokes-on-women-indians-netizens-left-guessing-why-11676171036353.html>). *Livemint*. livemint.com. Retrieved March 6, 2023.
121. Mitchell, Alex (February 14, 2023). "Great — now 'liberal' ChatGPT is censoring the[sic] The Post's Hunter Biden coverage, too" (<https://nypost.com/2023/02/14/chatgpt-censors-new-york-post-coverage-of-hunter-biden/>). *New York Post*. nypost.com. Retrieved February 6, 2023.

122. Guynn, Jessica. "Is ChatGPT 'woke'? AI chatbot accused of anti-conservative bias and a grudge against Trump" (<https://www.usatoday.com/story/tech/2023/02/09/woke-chatgpt-conservatives-bias/11215353002/>). *USA Today*. Retrieved March 1, 2023.
123. Bray, Hiawatha (February 9, 2023). "Is ChatGPT liberal or conservative? Depends who you ask. - The Boston Globe" (<https://www.bostonglobe.com/2023/02/09/business/are-chatbots-liberal-or-conservative-depends-who-you-ask/>). *Boston Globe*. Retrieved March 1, 2023.
124. Vincent, James (February 17, 2023). "As conservatives criticize 'woke AI,' here are CI culture war queries" (<https://www.theverge.com/2023/2/17/23603906/openai-chatgpt->  
*The Verge*. Retrieved March 1, 2023.
125. Schechner, Sam; Kruppa, Miles (February 10, 2023). "Google Opens ChatGPT Rival Bard for Testing, as Up" (<https://www.wsj.com/articles/google-opens-testing-of-chatgpt-rival-as-artificial-intelligence-war-heats-up-11675711198>). *The Wall Street Journal*. Archived (<https://web.archive.org/web/20230206205031/https://www.wsj.com/articles/google-opens-testing-of-chatgpt-rival-as-artificial-intelligence-war-heats-up-11675711198>) from the original on February 6, 2023. Retrieved February 6, 2023.
126. Ray, Tiernan (January 23, 2023). "ChatGPT is 'not particularly innovative,' and 'nothing revolutionary,' says Meta's chief AI scientist" (<https://www.zdnet.com/article/chatgpt-is-not-particularly-innovative-and-nothing-revolutionary-says-metas-chief-ai-scientist/>). *ZDNET*. Retrieved February 16, 2023.
127. "Introducing LLaMA: A foundational, 65-billion-parameter language model" (<https://ai.facebook.com/blog/large-language-model-llama-meta-ai/>). *ai.facebook.com*. Retrieved March 3, 2023.
128. Toh, Michelle (February 7, 2023). "Baidu stock surges after announcement of ChatGPT-style AI bot" (<https://www.cnn.com/2023/02/06/tech/china-baidu-ai-bot-chatgpt-rival-intl-hnk/index.html>). *CNN*. Archived (<https://web.archive.org/web/20230208163313/https://www.cnn.com/2023/02/06/tech/china-baidu-ai-bot-chatgpt-rival-intl-hnk/index.html>) from the original on February 8, 2023. Retrieved February 8, 2023.
129. Jo He-rim (February 3, 2023). "Naver to introduce search GPT in first half of year" (<https://m.koreaherald.com/view.php?ud=20230203000553>). *The Korea Herald*.
130. "Yandex plans to develop alternative to ChatGPT neural network" (<https://tass.com/economy/1570119>). *Tass*. February 1, 2023.

## External links

- Official website (<https://chat.openai.com/chat>) 
- White paper (<https://arxiv.org/abs/2203.02155>) for InstructGPT, ChatGPT's predecessor

Retrieved from "<https://en.wikipedia.org/w/index.php?title=ChatGPT&oldid=1143497576>"

■