

GPT-4 Cheat Sheet: What Is GPT-4, and What Is it Capable Of?



by **Megan Crouse** in **Artificial Intelligence**
on September 1, 2023, 1:43 PM EDT

How much better is GPT-4 compared to previous models?
Learn about cost and capabilities.

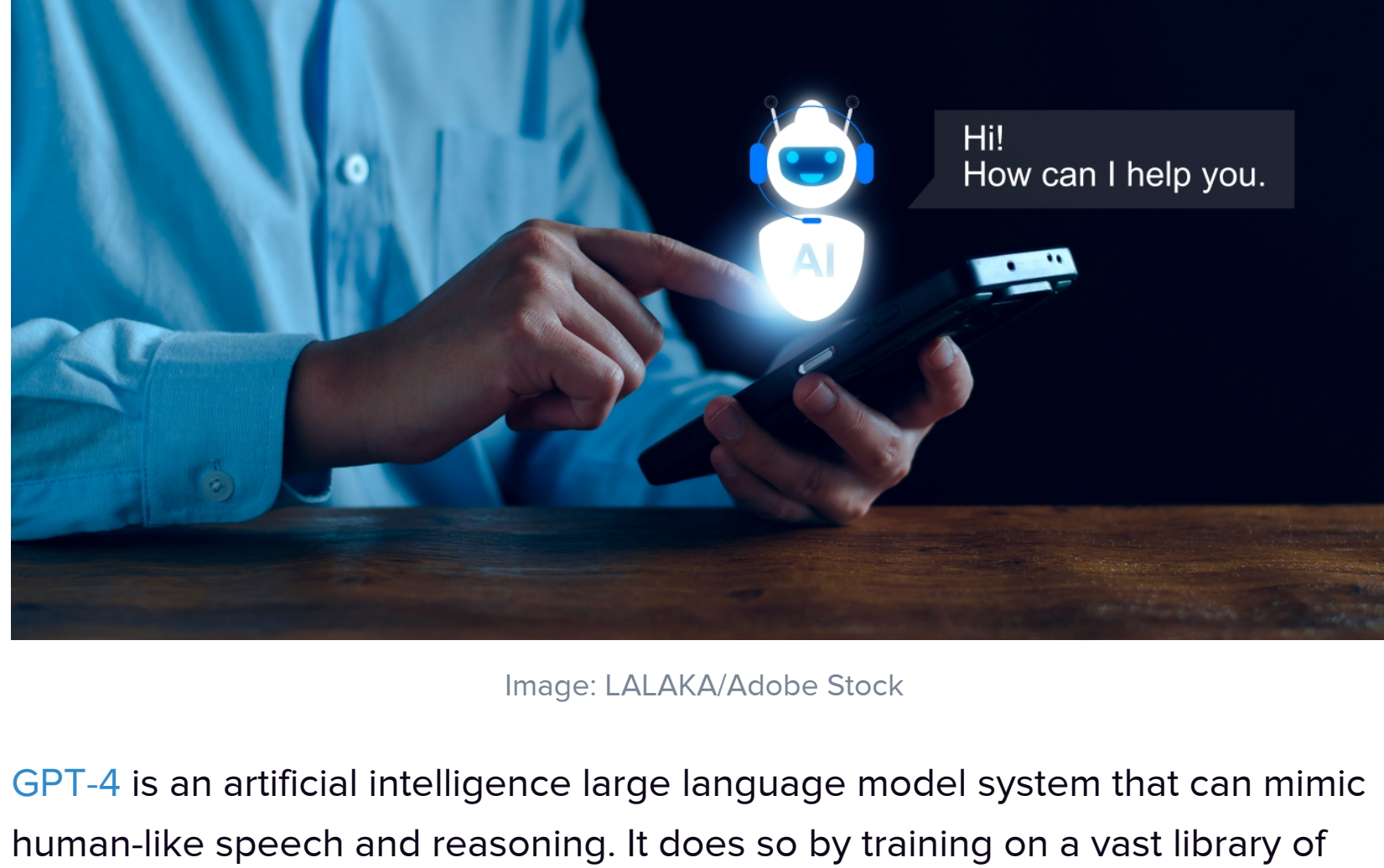


Image: L.LAKA/Adobe Stock

GPT-4 is an artificial intelligence large language model system that can mimic human-like speech and reasoning. It does so by training on a vast library of existing human communication, from classic works of literature to large swaths of the internet.

Artificial intelligence of this type builds on its training to predict what letter, number or other character is likely to come in sequence. This cheat sheet explores GPT-4 from a high level: how to access GPT-4 for either consumer or business use, who made it and how it works.

Jump to:

- [What is GPT-4?](#)
- [Who owns GPT-4?](#)
- [When was GPT-4 released?](#)
- [How can you access GPT-4?](#)
- [How much does GPT-4 cost to use?](#)
- [What are the capabilities of GPT-4?](#)
- [What is Bing Chat?](#)
- [What are the limitations of GPT-4 for business?](#)
- [GPT-4 vs. GPT-3.5 or ChatGPT](#)
- [Is upgrading to GPT-4 worth it?](#)
- [The latest GPT-4 trends](#)

What is GPT-4?

GPT-4 is a large multimodal model that can mimic prose, art, video or audio produced by a human. GPT-4 is able to solve written problems or generate original text or images. GPT-4 is the fourth generation of [OpenAI's foundation model](#).

The GPT-4 API, as well as the GPT-3.5 Turbo, DALL-E, and Whisper APIs, are now generally available as of July 7, 2023.

Who owns GPT-4?

GPT-4 is owned by OpenAI, an independent artificial intelligence company based in San Francisco. OpenAI was founded in 2015; it started out as a nonprofit but has since shifted to a for-profit model. OpenAI has received funding from Elon Musk, Microsoft, Amazon Web Services, Infosys, and other corporate and individual backers.

OpenAI has also produced [ChatGPT](#), a free-to-use chatbot spun out of the previous generation model, GPT-3.5, and DALL-E, an image-generating deep learning model. As the technology improves and grows in its capabilities, OpenAI reveals less and less about how its AI solutions are trained.

When was GPT-4 released?

[OpenAI announced its release of GPT-4 on March 14, 2023](#). It was immediately available for ChatGPT Plus subscribers, while other interested users needed to join a waitlist for access.

SEE: [Salesforce looped generative AI into its sales and field service products](#).

How can you access GPT-4?

The public version of GPT-4 is available at the [ChatGPT portal site](#).

On July 7, 2023, OpenAI made [the GPT-4 API](#) available for general use for "all existing API developers with a history of successful payments." OpenAI also expects to open access to new developers by the end of July. Rate-limits may be raised after that period depending on the amount of compute resources available.

In August 2023, GPT-4 was packaged as part of [ChatGPT Enterprise](#). Users of the business-oriented subscription receive unlimited use of a high-speed pipeline to GPT-4.

How much does GPT-4 cost to use?

For an individual, the ChatGPT Plus subscription costs \$20 per month to use.

Pricing for the text-only GPT-4 API starts at \$0.03 per 1k prompt tokens (one token is about four characters in English) and \$0.06 per 1k completion (output) tokens, OpenAI said. (OpenAI explains more about how tokens are counted [here](#).)

SEE: [Artificial Intelligence Ethics Policy \(TechRepublic Premium\)](#)

A second option with greater context length – about 50 pages of text – known as gpt-4-32k is also available. This option costs \$0.06 per 1K prompt tokens and \$0.12 per 1k completion tokens.

Other AI assistance services like GitHub's Copilot X now integrate GPT-4.

What are the capabilities of GPT-4?

Like its predecessor, GPT-3.5, GPT-4's main claim to fame is its output in response to natural language questions and other prompts. [OpenAI says](#) GPT-4 can "follow complex instructions in natural language and solve difficult problems with accuracy." Specifically, GPT-4 can solve math problems, answer questions, make inferences or tell stories. In addition, GPT-4 can summarize large chunks of content, which could be useful for either consumer reference or business use cases, such as a nurse summarizing the results of their visit to a client.

More must-read AI coverage

- [ChatGPT cheat sheet: Complete guide for 2023](#)
- [Generative AI Defined: How It Works, Benefits and Dangers](#)
- [Microsoft's First Generative AI Certificate Is Available for Free](#)
- [Hiring kit: Prompt engineer](#)

OpenAI tested GPT-4's ability to repeat information in a coherent order using several skills assessments, including AP and Olympiad exams and the Uniform Bar Examination. It scored in the 90th percentile on the Bar Exam and the 93rd percentile on the SAT Evidence-Based Reading & Writing exam. GPT-4 earned varying scores on AP exams.

These are not true tests of knowledge; instead, running GPT-4 through standardized tests shows the model's ability to form correct-sounding answers out of the mass of preexisting writing and art it was trained on.

GPT-4 predicts which token is likely to come next in a sequence. (One token may be a section of a string of numbers, letters, spaces or other characters.) While OpenAI is closed-mouthed about the specifics of GPT-4's training, LLMs are typically trained by first translating information in a dataset into tokens; the dataset is then cleaned to remove garbled or repetitive data. Next, AI companies typically employ people to apply reinforcement learning to the model, nudging the model toward responses that make common sense. The weights, which put very simply are the parameters that tell the AI which concepts are related to each other, may be adjusted in this stage.

The Chat Completions API and its upgrades

The [Chat Completions API](#) lets developers use the GPT-4 API through a freeform text prompt format. With it, they can build chatbots or other functions requiring back-and-forth conversation. The Chat Completions API first became available in June 2020.

In January 2024, the Chat Completions API will be upgraded to use newer completion models. OpenAI's ada, babbage, curie, and davinci models will be upgraded to version 002, while Chat Completions tasks using other models will transition to gpt-3.5-turbo-instruct.

GPT-3.5 Turbo fine-tuning and other news

On August 22 OpenAI announced the availability of [fine-tuning for GPT-3.5 Turbo](#). This enables developers to customize models and test those custom models for their specific use cases.

OpenAI expects fine-tuning to come to GPT-4 in fall 2023.

What is Bing Chat?

Microsoft's Bing Chat is an AI assistant deployed as a sidebar alongside the search engine Bing. Users can ask it to answer questions or generate images. Bing Chat runs on GPT-4. It sets itself apart from ChatGPT by being able to [remember previous conversations](#), although sometimes which chatbot can do this [doesn't match expectations](#).

SEE: [Discover to query Bing Chat to get the results you want. \(TechRepublic\)](#)

Bing Chat requires a Microsoft and the Edge browser to use.

In addition, Microsoft offers [Bing Chat Enterprise](#), which adds data protections and additional functionality to Bing Chat. Admins can provide managed access to Bing Chat Enterprise through Microsoft Entra ID (Azure Active Directory).

What are the limitations of GPT-4 for business?

Like other AI tools of its ilk, GPT-4 has limitations. For example, GPT-4 does not check if its statements are accurate. Its training on text and images from throughout the internet can make its responses nonsensical or inflammatory. However, OpenAI has digital controls and human trainers to try to keep the output as useful and business-appropriate as possible.

Additionally, GPT-4 tends to create 'hallucinations,' which is the artificial intelligence term for inaccuracies. Its words may make sense in sequence since they're based on probabilities established by what the system was trained on, but they aren't fact-checked or directly connected to real events. OpenAI is working on reducing the number of falsehoods the model produces.

Another major limitation is the question of whether sensitive corporate information that's fed into GPT-4 will be used to train the model and expose that data to external parties. Microsoft, which has a resale deal with OpenAI, plans to offer private ChatGPT instances to corporations later in the second quarter of 2023, according to [an April report](#).

Like GPT-3.5, GPT-4 does not incorporate information more recent than September 2021 in its lexicon. One of GPT-4's competitors, Google Bard, does have up-to-the-minute information because it is trained on the contemporary internet.

AI can suffer model collapse when trained on AI-created data; this problem is becoming more common as AI models proliferate.

GPT-4 vs. GPT-3.5 or ChatGPT

OpenAI's second most recent model, GPT-3.5, differs from the current generation in a few ways. OpenAI has not revealed the size of the model that GPT-4 was trained on but says it is "more data and more computation" than the billions of parameters ChatGPT was trained on. GPT-4 has also shown more deftness when it comes to writing a wider variety of materials, including fiction.

GPT-4 performs higher than ChatGPT on the standardized tests mentioned above. Answers to prompts given to the chatbot may be more concise and easier to parse. OpenAI notes that GPT-3.5 Turbo matches or outperforms GPT-4 on [certain custom tasks](#).

Additionally, GPT-4 is better than GPT-3.5 at making business decisions, such as scheduling or summarization. GPT-4 is "82% less likely to respond to requests for disallowed content and 40% more likely to produce factual responses," [OpenAI said](#).

SEE: [Learn how to use ChatGPT. \(TechRepublic Academy\)](#)

Another large difference between the two models is that [GPT-4 can handle images](#). It can serve as a visual aid, describing objects in the real world or determining the most important elements of a website and describing them.

"Over a range of domains — including documents with text and photographs, diagrams or screenshots — GPT-4 exhibits similar capabilities as it does on text-only inputs," OpenAI wrote in its [GPT-4 documentation](#).

Is upgrading to GPT-4 worth it?

Whether the new capabilities offered through GPT-4 are appropriate for your business is a decision that largely depends upon your use cases and whether you have found success with natural language artificial intelligence. Review the capabilities and limitations listed above, and consider where GPT-4 might be able to save time or reduce costs; conversely, consider which tasks might materially benefit from human knowledge, skill and common sense.

The latest GPT-4 trends

Microsoft announced in early August 2023 that GPT-4 availability in Azure OpenAI Service has expanded to [several new coverage regions](#).

Fine-tuning for GPT-4, which allows users to customize models, is expected to be available in the fall, OpenAI said.

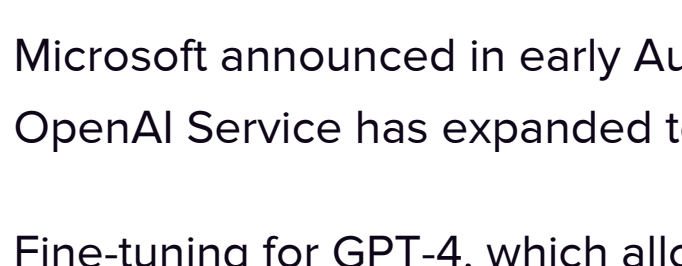
Subscribe to the TechRepublic Premium Exclusives Newsletter

Save time with the latest TechRepublic Premium downloads, including original research, customizable IT policy templates, ready-made lunch-and-learn presentations, IT hiring tools, ROI calculators, and more. Exclusively for you! Delivered Tuesdays and Thursdays

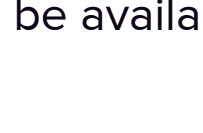
[✉ Sign up today](#)

Also See

- [ChatGPT cheat sheet: Complete guide for 2023](#) (TechRepublic)
- [Hiring Kit: Artificial Intelligence Architect](#) (TechRepublic Premium)
- [Hiring Kit: Machine Learning Engineer](#) (TechRepublic Premium)
- [Artificial Intelligence: More must-read coverage](#) (TechRepublic on Flipboard)



Share: GPT-4 Cheat Sheet: What Is GPT-4, and What Is it Capable Of?



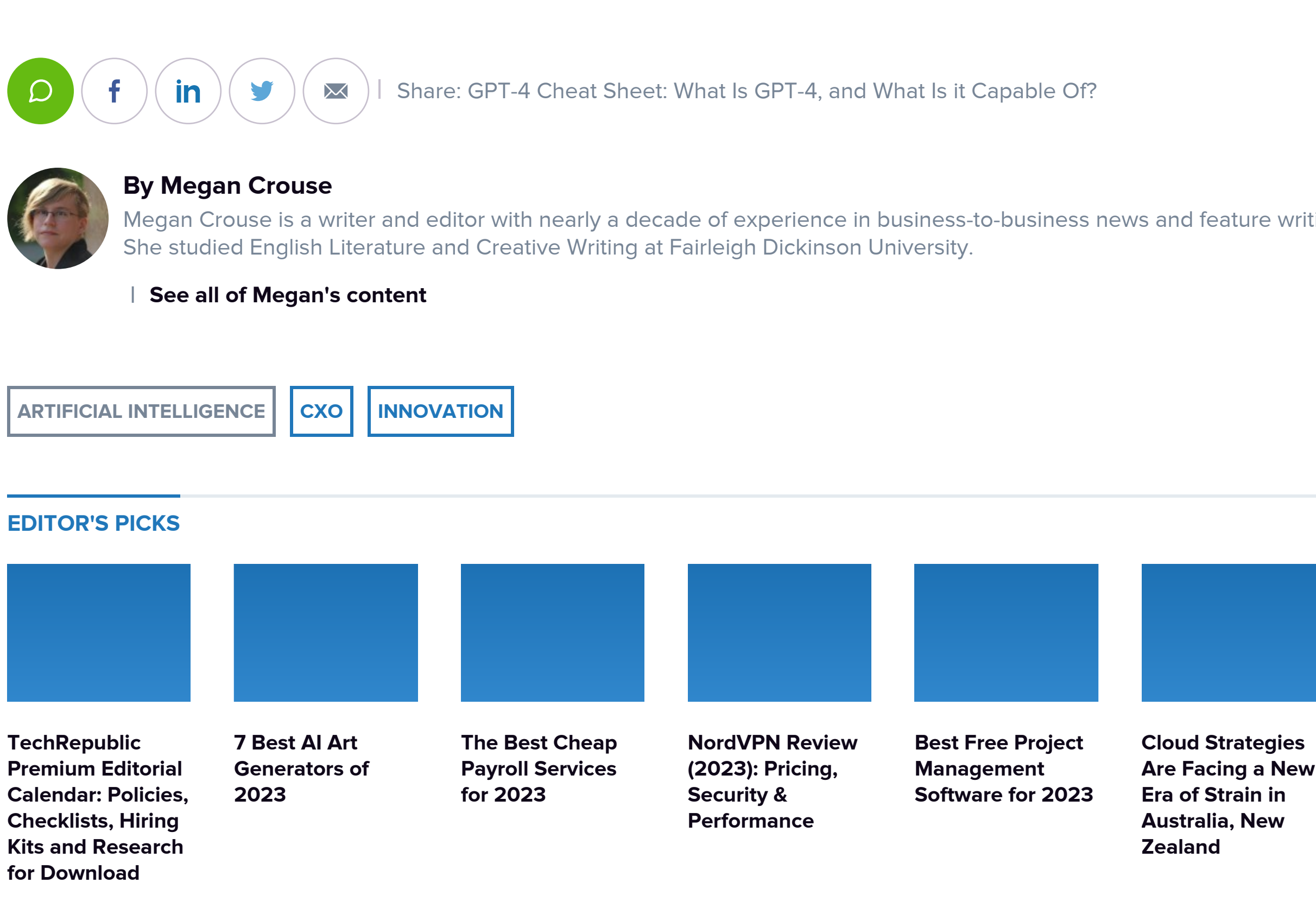
By **Megan Crouse**

Megan Crouse is a writer and editor with nearly a decade of experience in business-to-business news and feature writing. She studied English Literature and Creative Writing at Fairleigh Dickinson University.

[| See all of Megan's content](#)

[ARTIFICIAL INTELLIGENCE](#) [CXO](#) [INNOVATION](#)

EDITOR'S PICKS



TechRepublic Premium Editorial Calendar: Policies, Checklists, Hiring Kits and Research for Download

7 Best AI Art Generators of 2023

The Best Cheap Payroll Services for 2023

NordVPN Review (2023): Pricing, Security & Performance

Best Free Project Management Software for 2023

Cloud Strategies Are Facing a New Era of Strain in Australia, New Zealand



SERVICES

[Third-Party Vendor Policy](#)

[About Us](#) [Advertise](#)

[Newsletters](#) [Do Not Sell My Information](#)

[RSS Feeds](#) [Careers](#)

[Site Map](#)

[Site Help & Feedback](#)

[FAQ](#)

EXPLORE

[Downloads](#) [Resource Library](#)

[TechRepublic Forums](#) [Photos](#)

[Meet the Team](#) [Videos](#)

[TechRepublic Academy](#)

[TechRepublic Premium](#)