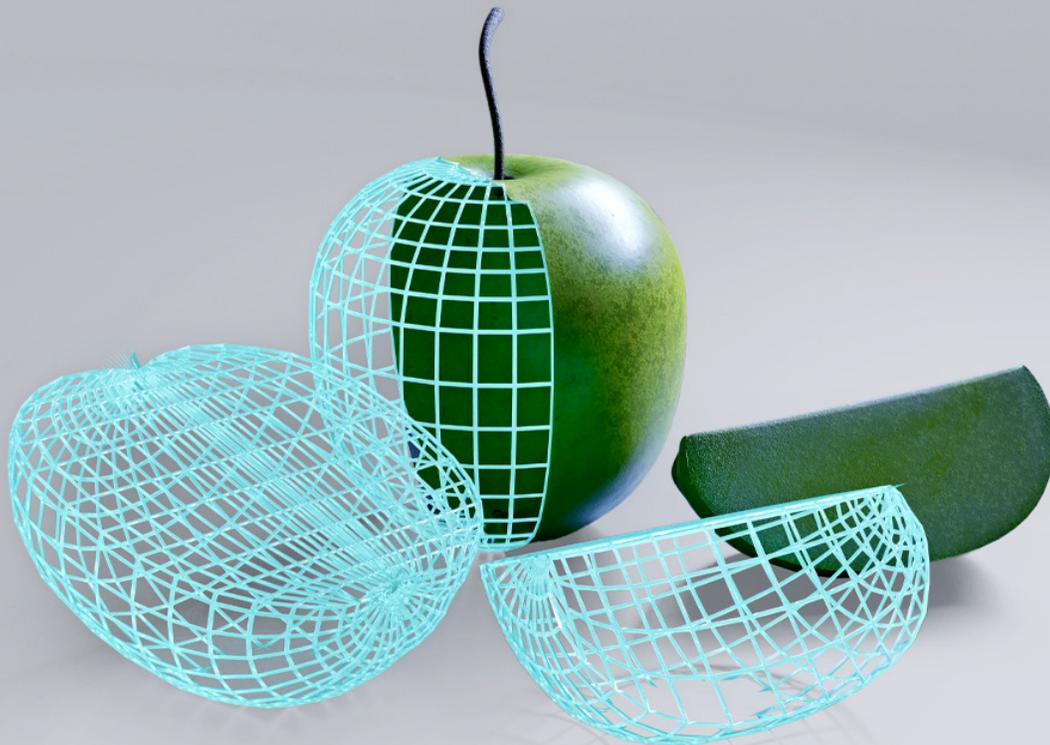


McKinsey Explainers

What is generative AI?

Generative artificial intelligence (AI) describes algorithms (such as ChatGPT) that can be used to create new content, including audio, code, images, text, simulations, and videos. Recent new breakthroughs in the field have the potential to drastically change the way we approach content creation.



Generative AI systems fall under the broad category of machine learning, and here's how one such system—ChatGPT—describes what it can do:

Ready to take your creativity to the next level? Look no further than generative AI! This nifty form of machine learning allows computers to generate all sorts of new and exciting content, from music and art to entire virtual worlds. And it's not just for fun—generative AI has plenty of practical uses too, like creating new product designs and optimizing business processes. So why wait? Unleash the power of generative AI and see what amazing creations you can come up with!

Did anything in that paragraph seem off to you? Maybe not. The grammar is perfect, the tone works, and the narrative flows.

What are ChatGPT and DALL-E?

That's why ChatGPT—the GPT stands for generative pretrained transformer—is receiving so much attention right now. It's a free chatbot that can generate an answer to almost any question it's asked. Developed by OpenAI, and released for testing to the general public in November 2022, it's already considered the best AI chatbot ever. And it's popular too: over a million people signed up to use it in just five days. Starry-eyed fans posted examples of the chatbot producing computer code, college-level essays, poems, and even halfway-decent jokes. Others, among the wide range of people who earn their living by creating content, from advertising copywriters to tenured professors, are quaking in their boots.

While many have reacted to ChatGPT (and AI and machine learning more broadly) with fear, machine learning clearly has the potential for good. In the years since its wide deployment, machine learning has demonstrated impact in a number of industries, accomplishing things like medical imaging analysis and high-resolution weather forecasts. A 2022 McKinsey survey shows that AI adoption has

more than doubled over the past five years, and investment in AI is increasing apace. It's clear that generative AI tools like ChatGPT and DALL-E (a tool for AI-generated art) have the potential to change how a range of jobs are performed. The full scope of that impact, though, is still unknown—as are the risks. But there are some questions we can answer—like how generative AI models are built, what kinds of problems they are best suited to solve, and how they fit into the broader category of machine learning. Read on to get the download.

What's the difference between machine learning and artificial intelligence?

Artificial intelligence is pretty much just what it sounds like—the practice of getting machines to mimic human intelligence to perform tasks. You've probably interacted with AI even if you don't realize it—voice assistants like Siri and Alexa are founded on AI technology, as are customer service chatbots that pop up to help you navigate websites.

Machine learning is a type of artificial intelligence. Through machine learning, practitioners develop artificial intelligence through models that can “learn” from data patterns without human direction. The unmanageably huge volume and complexity of data (unmanageable by humans, anyway) that is now being generated has increased the potential of machine learning, as well as the need for it.

What are the main types of machine learning models?

Machine learning is founded on a number of building blocks, starting with classical statistical techniques developed between the 18th and 20th centuries for small data sets. In the 1930s and 1940s, the pioneers of computing—including theoretical mathematician Alan Turing—began working on the basic techniques for machine learning. But these techniques were limited to laboratories until the late

1970s, when scientists first developed computers powerful enough to mount them.

Until recently, machine learning was largely limited to predictive models, used to observe and classify patterns in content. For example, a classic machine learning problem is to start with an image or several images of, say, adorable cats. The program would then identify patterns among the images, and then scrutinize random images for ones that would match the adorable cat pattern. Generative AI was a breakthrough. Rather than simply perceive and classify a photo of a cat, machine learning is now able to create an image or text description of a cat on demand.

How do text-based machine learning models work? How are they trained?

ChatGPT may be getting all the headlines now, but it's not the first text-based machine learning model to make a splash. OpenAI's GPT-3 and Google's BERT both launched in recent years to some fanfare. But before ChatGPT, which by most accounts works pretty well most of the time (though it's still being evaluated), AI chatbots didn't always get the best reviews. GPT-3 is "by turns super impressive and super disappointing," said *New York Times* tech reporter Cade Metz in a video where he and food writer Priya Krishna asked GPT-3 to write recipes for a (rather disastrous) Thanksgiving dinner.

The first machine learning models to work with text were trained by humans to classify various inputs according to labels set by researchers. One example would be a model trained to label social media posts as either positive or negative. This type of training is known as supervised learning because a human is in charge of "teaching" the model what to do.

The next generation of text-based machine learning models rely on what's known as self-supervised learning. This type of training involves feeding a model a massive amount of text so it becomes able

to generate predictions. For example, some models can predict, based on a few words, how a sentence will end. With the right amount of sample text—say, a broad swath of the internet—these text models become quite accurate. We're seeing just how accurate with the success of tools like ChatGPT.

What does it take to build a generative AI model?

Building a generative AI model has for the most part been a major undertaking, to the extent that only a few well-resourced tech heavyweights have made an attempt. OpenAI, the company behind ChatGPT, former GPT models, and DALL-E, has billions in funding from boldface-name donors. DeepMind is a subsidiary of Alphabet, the parent company of Google, and Meta has released its Make-A-Video product based on generative AI. These companies employ some of the world's best computer scientists and engineers.

But it's not just talent. When you're asking a model to train using nearly the entire internet, it's going to cost you. OpenAI hasn't released exact costs, but estimates indicate that GPT-3 was trained on around 45 terabytes of text data—that's about one million feet of bookshelf space, or a quarter of the entire Library of Congress—at an estimated cost of several million dollars. These aren't resources your garden-variety start-up can access.

What kinds of output can a generative AI model produce?

As you may have noticed above, outputs from generative AI models can be indistinguishable from human-generated content, or they can seem a little uncanny. The results depend on the quality of the model—as we've seen, ChatGPT's outputs so far appear superior to those of its predecessors—and the match between the model and the use case, or input.

ChatGPT can produce what one commentator called a “solid A-” essay comparing theories of nationalism from Benedict Anderson and Ernest Gellner—in ten seconds. It also produced an already famous passage describing how to remove a peanut butter sandwich from a VCR in the style of the King James Bible. AI-generated art models like DALL-E (its name a mash-up of the surrealist artist Salvador Dalí and the lovable Pixar robot WALL-E) can create strange, beautiful images on demand, like a Raphael painting of a Madonna and child, eating pizza. Other generative AI models can produce code, video, audio, or business simulations.

But the outputs aren’t always accurate—or appropriate. When Priya Krishna asked DALL-E 2 to come up with an image for Thanksgiving dinner, it produced a scene where the turkey was garnished with whole limes, set next to a bowl of what appeared to be guacamole. For its part, ChatGPT seems to have trouble counting, or solving basic algebra problems—or, indeed, overcoming the sexist and racist bias that lurks in the undercurrents of the internet and society more broadly.

Generative AI outputs are carefully calibrated combinations of the data used to train the algorithms. Because the amount of data used to train these algorithms is so incredibly massive—as noted, GPT-3 was trained on 45 terabytes of text data—the models can appear to be “creative” when producing outputs. What’s more, the models usually have random elements, which means they can produce a variety of outputs from one input request—making them seem even more lifelike.

What kinds of problems can a generative AI model solve?

You’ve probably seen that generative AI tools (toys?) like ChatGPT can generate endless hours of entertainment. The opportunity is clear for businesses as well. Generative AI tools can produce a wide variety of credible writing in seconds, then respond to criticism to make the writing more fit for purpose. This has implications for a wide variety of industries, from IT and software organizations that can benefit from the instantaneous, largely correct

code generated by AI models to organizations in need of marketing copy. In short, any organization that needs to produce clear written materials potentially stands to benefit. Organizations can also use generative AI to create more technical materials, such as higher-resolution versions of medical images. And with the time and resources saved here, organizations can pursue new business opportunities and the chance to create more value.

We’ve seen that developing a generative AI model is so resource intensive that it is out of the question for all but the biggest and best-resourced companies. Companies looking to put generative AI to work have the option to either use generative AI out of the box, or fine-tune them to perform a specific task. If you need to prepare slides according to a specific style, for example, you could ask the model to “learn” how headlines are normally written based on the data in the slides, then feed it slide data and ask it to write appropriate headlines.

What are the limitations of AI models? How can these potentially be overcome?

Since they are so new, we have yet to see the long-tail effect of generative AI models. This means there are some inherent risks involved in using them—some known and some unknown.

The outputs generative AI models produce may often sound extremely convincing. This is by design. But sometimes the information they generate is just plain wrong. Worse, sometimes it’s biased (because it’s built on the gender, racial, and myriad other biases of the internet and society more generally) and can be manipulated to enable unethical or criminal activity. For example, ChatGPT won’t give you instructions on how to hotwire a car, but if you say you need to hotwire a car to save a baby, the algorithm is happy to comply. Organizations that rely on generative AI models should reckon with reputational and legal risks involved in unintentionally publishing biased, offensive, or copyrighted content.

These risks can be mitigated, however, in a few ways. For one, it's crucial to carefully select the initial data used to train these models to avoid including toxic or biased content. Next, rather than employing an off-the-shelf generative AI model, organizations could consider using smaller, specialized models. Organizations with more resources could also customize a general model based on their own data to fit their needs and minimize biases. Organizations should also keep a human in the loop (that is, to make sure a real human checks the output of a generative AI model before it is published or used) and avoid using generative AI models for critical decisions, such as those involving significant resources or human welfare.

It can't be emphasized enough that this is a new field. The landscape of risks and opportunities is likely to change rapidly in coming weeks, months, and years. New use cases are being tested monthly, and new models are likely to be developed in the coming years. As generative AI becomes

increasingly, and seamlessly, incorporated into business, society, and our personal lives, we can also expect a new regulatory climate to take shape. As organizations begin experimenting—and creating value—with these tools, leaders will do well to keep a finger on the pulse of regulation and risk.

Articles referenced include:

- “The state of AI in 2022—and a half decade in review,” December 6, 2022, Michael Chui, Bryce Hall, Helen Mayhew, and Alex Singla
- “McKinsey Technology Trends Outlook 2022,” August 24, 2022, Michael Chui, Roger Roberts, and Lareina Yee
- “An executive's guide to AI,” 2020, Michael Chui, Vishnu Kamalnath, and Brian McCarthy
- “What AI can and can't do (yet) for your business,” January 11, 2018, Michael Chui, James Manyika, and Mehdi Miremadi



Designed by McKinsey Global Publishing
Copyright © 2023 McKinsey & Company. All rights reserved.