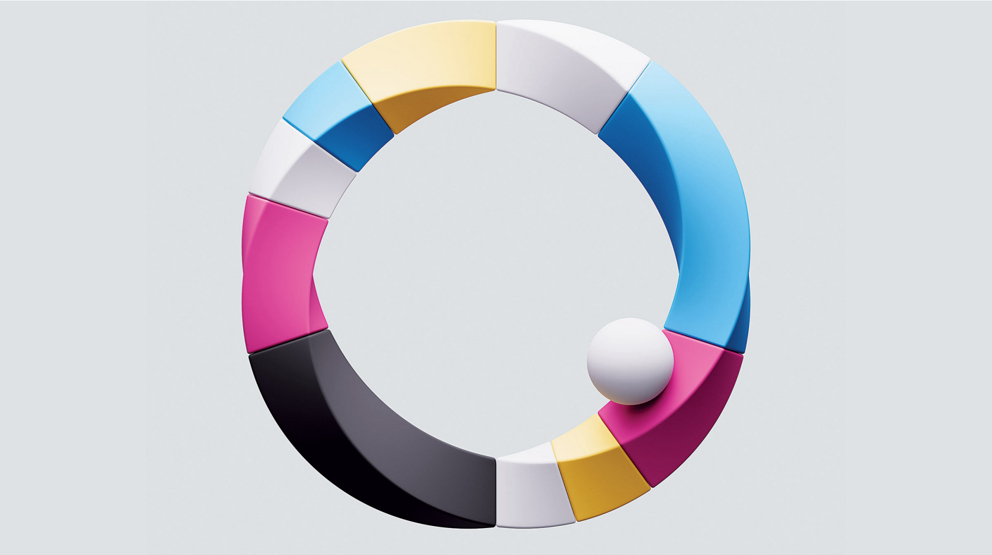
Harvard Business Review

Spotlight Article

AI and Machine Learning



How to Capitalize on Generative AI

A guide to realizing its benefits while limiting its risks

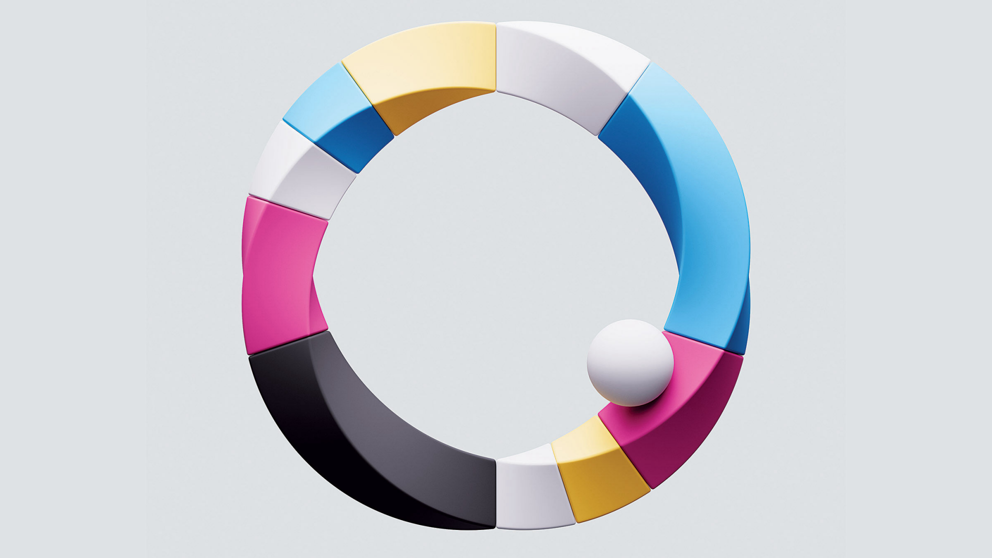
by Andrew McAfee, Daniel Rock, and Erik Brynjolfsson

How to Capitalize on Generative AI

A guide to realizing its benefits while limiting its risks

by Andrew McAfee, Daniel Rock, and Erik Brynjolfsson

From the Magazine (November–December 2023) / Reprint S23061



Michael Brandon Myers

Business leaders are struggling to understand how seriously they should take the latest phenomenon in the world of artificial intelligence: generative AI. On one hand, it has already displayed a breathtaking ability to create new content such as music, speech, text, images, and video and is currently used, for instance, to write software, to transcribe physicians' interactions with their patients, and to allow people to converse with a customer-relationship-management system. On the other hand, it is far from perfect: It sometimes produces distorted or entirely fabricated output and can be oblivious to privacy and copyright concerns.

Is generative AI's importance overblown? Are its risks worth the potential rewards? How can companies figure out where best to apply it? What should their first steps be? To provide guidance, this article draws on our research comprising studies of specific generative-AI projects and broad analyses of how the technology will affect tasks and jobs throughout the economy.

A large enterprise-software company that one of us (Erik) studied along with Lindsey Raymond and Danielle Li of MIT illustrates that there are ways to both reap the benefits of generative AI and contain its risks. The company's customer-service agents, who assist people via online chats, faced a common challenge: New hires needed several months to get up to speed on how to answer technical questions and deal with confused customers, but many quit before they became proficient. The company saw generative AI as a solution. It engaged Cresta (which Erik has been advising), a generative AI start-up, to implement two kinds of artificial intelligence. The first was a large language model (LLM)—designed to understand and respond to humans in their own words—that "listened in" on the chats. It was fine-tuned to recognize phrases that led to good customer-service outcomes in various situations. But because of the risk of *confabulations,* or plausible-sounding but incorrect responses, the system also used a machine-learning technique called *in-context learning,* which drew answers from relevant user manuals and documents.

The LLM monitored the online chats for specific phrases, and when one of them occurred, it based its responses on the information in the in-context learning system. As an additional safeguard, it didn't respond to queries directly. Instead human agents were free to apply their common sense in deciding whether to use or ignore the LLM's suggestions.

After a seven-week pilot the system was rolled out to more than 1,500 agents. Within two months multiple benefits appeared. Both the average number of issues resolved per hour and the number of chats an agent could handle simultaneously increased by almost 15%; the average chat time decreased by nearly 10%; and an analysis of the chat logs showed that immediately after the new system was implemented, customer satisfaction improved. Expressions of frustration declined, for example, as did TYPING IN ALL CAPS.

It's especially interesting that the least-skilled agents, who were also often the newest, benefited most. For example, resolutions per hour by agents who had been among the slowest 20% before introduction of the new system increased by 35%. (The resolution rate of the fastest 20% didn't change.) The generative AI system was a fast-acting upskilling technology. It made available to all agents knowledge that had previously come only with experience or training. What's more, agent turnover fell, especially among those with less than six months of experience—perhaps because people are more likely to stick around when they have powerful tools to help them do their jobs better.

Given the potential of generative AI to improve productivity in many other functions—indeed, any that involve cognitive tasks—calling it revolutionary is no hyperbole. Business leaders should view it as a general-purpose technology akin to electricity, the steam engine, and the internet. But although the full potential of those other technologies took decades to be realized, generative AI's impact on performance and competition throughout the economy will be clear in just a few years.

That's because general-purpose technologies of the past required a great deal of complementary physical infrastructure (power lines, new kinds of motors and appliances, redesigned factories, and so on) along with new skills and business processes. That's not the case with generative AI. Much of the necessary infrastructure is already in place: The cloud, software-as-a-service, application programming interfaces, app stores, and other advances keep lowering the amount of time, effort, expertise, and expense needed to acquire and start using new information systems. As a result, it keeps getting easier for companies to deploy just about *any* digital technology. That's a big reason ChatGPT went from zero to 100 million users in 60 days. As Microsoft, Google, and other technology providers incorporate generative AI tools in their office suites, email clients, and other applications, billions of users will speedily gain access as part of their daily routine.

Generative AI will also deploy quickly because people interact with these systems by talking to them much as they would to another person. That lowers the barriers to entry for some kinds of work (imagine writing software by explaining to an LLM in everyday speech what you want to accomplish). In addition, these systems won't necessarily require companies to change entire business processes; at first they will be used for discrete tasks only, which will make them much easier to adopt. Using technology to reengineer every aspect of how a company interacts with its customers, for example, is a major undertaking; using it to suggest better chat responses to customer service agents is not. Over time, however, generative AI will bring large and deep changes in how companies do their most important work.

Consequently, business leaders shouldn't sit on the sidelines and wait to see how the use of generative AI develops. They can't afford to let competitors steal a march on them.

How Will Generative AI Affect Your Company's Jobs?

Predictions of the kinds and numbers of jobs that will be replaced by generative AI abound. But it's actually more helpful to think about the cognitive tasks that the technology could perform or help perform.

Research conducted by one of us (Daniel), OpenResearch's Sam Manning, and OpenAI's Tyna Eloundou and Pamela Mishkin took that approach. Their starting point was the O*NET database, which has been maintained and updated by the U.S. government since 1998. O*NET includes nearly 1,000 occupations and breaks each one down into its constituent tasks—typically 20 to 30 of them. For instance, according to O*NET, radiologists have 30 distinct tasks, including "perform or interpret the outcomes of diagnostic imaging procedures" and "develop treatment plans for radiology patients."

The researchers, with the assistance of people chosen by OpenAI, addressed two questions: Which tasks of each O*NET job could be done at least twice as fast with the help of generative AI with no significant drop in quality? And of those "exposed" tasks, which needed at least one system in addition to generative AI to reap the productivity gains? The research team also asked OpenAI's GPT-4 LLM the same two questions and compared its answers with those of the people. The answers were similar.

> **Calling generative AI revolutionary is no hyperbole. Business leaders should view it as a general-purpose technology akin to electricity, the steam engine, and the internet.**

This effort revealed that 80% of U.S. workers have at least 10% of their tasks exposed to generative AI, and 19% of workers have more than half of their tasks exposed. But "exposed" doesn't mean that those tasks will or should be automated. In many cases the best use of generative AI will be to make human workers more productive or creative, not to replace them. Programmers are a case in point. They already heavily use LLMs like GitHub Copilot to write the first draft of their code, but they still have to correct errors; consult with managerial, engineering, and technical personnel to clarify the program's intent; train subordinates; and perform many other tasks that are unsuitable for generative AI. As LLMs get better at writing code, programmers will have more time and energy to devote to other tasks. (For more about how generative AI can help but not replace workers, see "How Generative AI Can Augment Human Creativity," HBR, July–August 2023.)

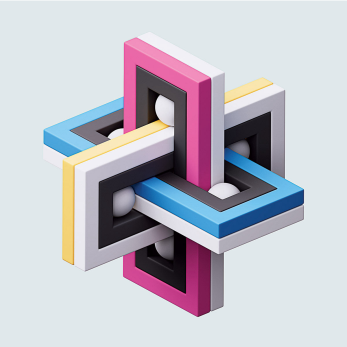
Leaders can undertake a version of this research approach to get a sense of where generative AI might be most productively applied in their organizations. Every board should expect its CEO to develop an actionable game plan. Doing so is a three-part process.

First, do a rough inventory of knowledge-work jobs: How many of your people primarily write for a living? How many data analysts, managers, programmers, customer service agents, and so on do you have?

Next, ask two questions about each role. The first is, "How much would an employee in this role benefit from having a competent but naive assistant—someone who excels at programming, writing, preparing data, or summarizing information but knows nothing about our company?" Today's publicly available LLMs are like such an assistant. They can write code, for example, but they don't know what your software development or systems integration needs are. They can create a project plan or critique an existing one, but they don't know what projects you're working on.

The second question is, "How much would an employee in this role benefit from having an experienced assistant—someone who's been at the company long enough to absorb its specialized knowledge?" The software company described at the start of this article didn't need naive customer-service agents; it needed agents who knew what kinds of problems occurred with its products and could work effectively with customers to solve them. That's why it combined a customer-facing LLM with in-context learning. As this example indicates, when a company needs access to its specific internal knowledge, it must typically combine "off the shelf" generative AI with another system.

Finally, once your company's knowledge-work roles have been inventoried and those two questions have been answered, prioritize the most-promising generative-AI efforts. This task is straightforward: Choose the ones with the largest benefit-to-cost ratio. To estimate benefits, look at the total amount the company is spending on compensation for each role. The purpose is not to identify positions for elimination; rather, it's to identify opportunities for big productivity improvements—where new digital assistants will be most valuable.



Michael Brandon Myers

As is the case with other digital-transformation efforts, the cost of a generative AI project is a combination of money, time, and lost opportunities—the projects you're not pursuing because generative AI is a higher priority. Off-the-shelf LLM efforts are relatively cheap and fast, whereas projects that require integrating generative AI with another system take longer and are more expensive (although not by comparison with many other IT efforts).

At present most generative AI projects are focused on improving specific tasks. That's appropriate, because so many opportunities exist to use the technology that way. But as it matures and companies gain experience with it, generative AI efforts will encompass entire business processes instead of individual tasks. For example, they'll be used to transform every aspect of a company's interaction with customers, not just to improve online troubleshooting chats. Generative AI is still a nascent technology, and we can't predict exactly how it will be put to work in the years ahead. But we can confidently predict that it will have a large role in the digital strategies of successful companies.

Remedying the "Confabulation" Problem

Given the major impact that generative AI promises to have on a wide variety of businesses in the near future, the response to one of its biggest shortcomings—that it can fabricate information—shouldn't be to avoid the technology. Rather, it should be to safeguard against that danger. Here are ways to do so.

Build multilevel LLMs or combine one with another system.
Companies that build LLMs are well aware that these systems confabulate and are working on ways to minimize the problem. One technique is to recognize when a user's request is not suitable for an LLM's standard approach, which is to formulate an answer on the basis of associations among all the words and sentences it has been trained on. For such requests, the system takes a different tack. For example, in response to a query that has a single right answer, Google's Bard now actually writes an algorithm to produce that answer, which it reports to the user (along with the code). For instance, when asked to reverse the word "Lollipop," it wrote a few lines of code to accomplish the task

and returned “popilloL.” And as noted, the project to improve customer service employed an LLM to monitor online chats and understand customer questions, but the in-context learning system shaped its responses.

Supplement the LLM with a human. Users should take an LLM’s output with a grain of salt. For example, marketers using an LLM to generate copy for a website or a social media campaign can look at what the system comes up with and quickly assess whether it’s on target. Software engineers can see whether the code produced by the generative AI runs and accomplishes the desired task. Even if it doesn’t, according to the engineers, the approach it uses can help them tackle the task at hand. And physicians using LLMs to transcribe and summarize visits with consenting patients are reporting major reductions in the time required to document those visits. One doctor told Steve Lohr of the *New York Times* that an LLM had cut the time he spent per day on this task from as much as two hours to 20 minutes or so. Doctors still have to review the AI-generated summaries, but they no longer have to simultaneously interact with their patients and try to take notes about the interaction. As a result, another doctor told Lohr, “AI has allowed me, as a physician, to be 100% present for my patients.” Similarly, in the customer service example the agents’ own judgment vetted the reasonableness of the AI’s answers.

Don’t use an LLM. Some tasks are too risky for generative AI to be involved at all. For example, a system that prescribes exactly the right medications 90% of the time but confabulates in one case out of 10 is unacceptably unsafe to be used on its own. It also wouldn’t save physicians any time, because they’d have to carefully check all its recommendations before passing them on to patients. Even for tasks in which safety is not an issue, the tendency of LLMs to confabulate can rule them out. When one of us (Andy) was putting together the endnotes for his most recent book, he was thrilled to learn that ChatGPT could take a list of books, articles, and websites and generate a set of properly formatted references for them. But upon checking its output, he was dismayed to find that some of the references were wrong. When he gave it the URL of an article, it sometimes generated a reference with a plausible but made-up title, gave an incorrect date of publication, or attributed the article to the wrong author. He found it quicker to create all the references by hand than to check every aspect of the ones generated by the LLM.

Mitigating Invasion of Privacy, Intellectual-Property Problems, and Bias

If you use a confidential report to help train a generative AI system, bits of the report's contents might later show up in the response to a prompt from someone who shouldn't have access to that information. Consequently, it's important to be clear on the privacy policies of any generative AI you're using. The good news is that LLMs and strict privacy are not at all incompatible. The Mayo Clinic, for example, has announced an effort to deploy an internal LLM that will help its health care providers search for information across sources including web pages, internal documents, and patient records. If a doctor requests, "Show me today's test results for all my patients," the LLM will generate queries to the electronic-health-records system and present the results. To make the system compliant with requirements of the Health Insurance Portability and Accountability Act (HIPAA) regarding the privacy and confidentiality of patient information, Mayo will designate which of its employees are authorized to access protected health information.

The response to one of generative AI's biggest shortcomings—that it can fabricate information—should be not to avoid the technology but to safeguard against that danger.

In addition to confabulations and privacy concerns, a risk with some LLMs is violation of intellectual property (IP) rights. ChatGPT has been trained on enormous amounts of text, some of which is still covered by copyright or other IP rights. The same is true of new image-generating AI systems like Stable Diffusion and Midjourney, both of which have been sued for copyright infringement. Companies may be exposed to legal liability if the generative-AI-produced images they use are found to be in violation of IP laws (see "Generative AI Has an Intellectual Property Problem," hbr.org, April 7, 2023). As a result, many organizations are waiting to see how court cases are decided before diving into generative AI. But to encourage immediate adoption, some creators of these systems are shielding customers from IP risk. Adobe, for example, has announced that it will indemnify users of Firefly, its image-generating AI (which was not trained on copyrighted images), against legal claims.

One final concern with generative AI, as with most other types of artificial intelligence, is bias. "Garbage in, garbage out" is one of the oldest sayings of the computer era, and it's true now more than ever. If a machine-learning system is trained on biased data, the results it generates will reflect that bias. If, for example, a company has hired only college graduates as programmers and uses its employment history to train a system that helps make hiring decisions, that system will most likely reject highly qualified coders who didn't go to or finish college. So be vigilant as you're putting generative AI to work. Ask yourself, "Are we hoping this system will provide results that are less biased than the data it's been trained on?" If the answer is yes, rethink the project.

Be Ready to Experiment

Over the past few decades leading organizations have employed the agile method for successfully developing and adopting new information systems (see "Embracing Agile," HBR, May 2016). They manage their efforts with repeated trials rather than extensive planning. They break projects up into short cycles that can be completed in a week or two, sometimes even less. Project-team members track progress and reflect on what they've learned before starting the next cycle. Often, in fact, the whole cycle is an experiment: The goal isn't so much to build something as to test a hypothesis and gain understanding.

Generative AI is ideally suited to this iterative approach. Its strengths and weaknesses are unlike those of any earlier systems. You must figure out how to phrase your prompts to get the most useful responses. You also frequently have to tell the system to try again and give it notes on how to do better. Asking it to assume a persona or directing it to change its tone or style is often effective. Interacting with an LLM in this manner is called "prompt engineering"—a young discipline that is still more art than science. So is figuring out how to prevent confabulations. The best way to begin learning these arts is to find a project with an attractive benefit-to-cost ratio and low risks and start trying things. The same approach should be used with more-ambitious efforts to work with generative AI, such as combining an LLM with other technologies. Rapid iteration is the best way to learn and make progress. The faster an organization can move through repeated OODA loops of *observing* the situation, *orienting* for action, *deciding* what to do, and then *acting,* the more it will learn, and the faster productivity gains and other benefits will appear.

• • •

Generative AI promises to have a major impact on how businesses operate—and within a few years, not decades from now. Its tendency to confabulate and its privacy, intellectual property, and bias risks are all legitimate concerns, but they can be contained. Leaders cannot afford to take a wait-and-see attitude. They should start exploring the technology's potential now.

A version of this article appeared in the November–December 2023 issue of Harvard Business Review.

Andrew McAfee is the cofounder and co-director of the MIT Initiative on the Digital Economy and the inaugural visiting fellow in Google's Technology and Society group. He is the author of the new book, *The Geek Way*, and coauthor with Erik Brynjoflsson of *The Second Machine Age*.

DR

Daniel Rock is an assistant professor of operations, information, and decisions at the University of Pennsylvania's Wharton School and a cofounder of Workhelix, which creates generative AI strategies and implementation plans for companies.

Erik Brynjolfsson is the director of the Stanford Digital Economy Lab, a professor at the Stanford Institute for Human-Centered AI, a research associate at the National Bureau of Economic Research, and a cofounder of Workhelix, which creates generative AI strategies and implementation plans for companies.