



McKinsey Explainers

What is AI (artificial intelligence)?

Artificial intelligence is a machine's ability to perform some cognitive functions we usually associate with human minds.



Humans and machines: a match made in [productivity](#) heaven. Our species wouldn't have gotten very far without our mechanized workhorses. From the wheel that revolutionized agriculture to the screw that held together increasingly complex construction projects to the robot-enabled assembly lines of today, machines have made life as we know it possible. And yet, despite their seemingly endless utility, humans have long feared machines—more specifically, the possibility that machines might someday [acquire human intelligence](#) and strike out on their own.

But we tend to view the possibility of sentient machines with fascination as well as fear. This curiosity has helped turn science fiction into actual science. Twentieth-century theoreticians, like computer scientist and mathematician Alan Turing, envisioned a future where machines could perform functions faster than humans. The work of Turing and others soon made this a reality. Personal calculators became widely available in the 1970s, and by 2016, the US census showed that [89 percent of American households](#) had a computer. Machines—*smart* machines at that—are now just an ordinary part of our lives and culture.

Those smart machines are also getting faster and more complex. Some computers have now crossed the [exascale](#) threshold, meaning they can perform as many calculations in a single second as an

individual could in [31,688,765,000 years](#). And beyond computation, which machines have long been faster at than we have, computers and other devices are now acquiring skills and perception that were once unique to humans and a few other species.

AI is a machine's ability to perform the cognitive functions we associate with human minds, such as perceiving, reasoning, learning, interacting with the environment, problem-solving, and even exercising creativity. 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 some customer service chatbots that pop up to help you navigate websites.

[Applied AI](#)—simply, artificial intelligence applied to real-world problems—has serious implications for the business world. By using artificial intelligence, companies have the potential to make business more efficient and profitable. But ultimately, the value of AI isn't in the systems themselves. Rather, it's in how companies use these systems to assist humans—and their ability to [explain](#) to shareholders and the public what these systems do—in a way that builds trust and confidence.

For more about AI, its history, its future, and how to apply it in business, read on.

Learn more about [QuantumBlack, AI by McKinsey](#).

Beyond computation, computers and other devices are now acquiring skills and perception that were once unique to humans and a few other species.

What is machine learning?

Machine learning is a form of artificial intelligence that can adapt to a wide range of inputs, including large sets of historical data, synthesized data, or human inputs. (Some machine learning algorithms are specialized in training themselves to detect patterns; this is called deep learning. See Exhibit 1.) These algorithms can detect patterns and learn how to make predictions and recommendations by processing data, rather than by receiving explicit programming instruction. Some algorithms can also adapt in response to new data and experiences to improve over time.

The volume and complexity of data that is now being generated, too vast for humans to process and apply efficiently, has increased the potential of machine learning, as well as the need for it. In the years since its widespread deployment, which began in the 1970s, machine learning has had an impact on a number of industries, including achievements in [medical-imaging analysis](#) and high-resolution weather forecasting.

What is deep learning?

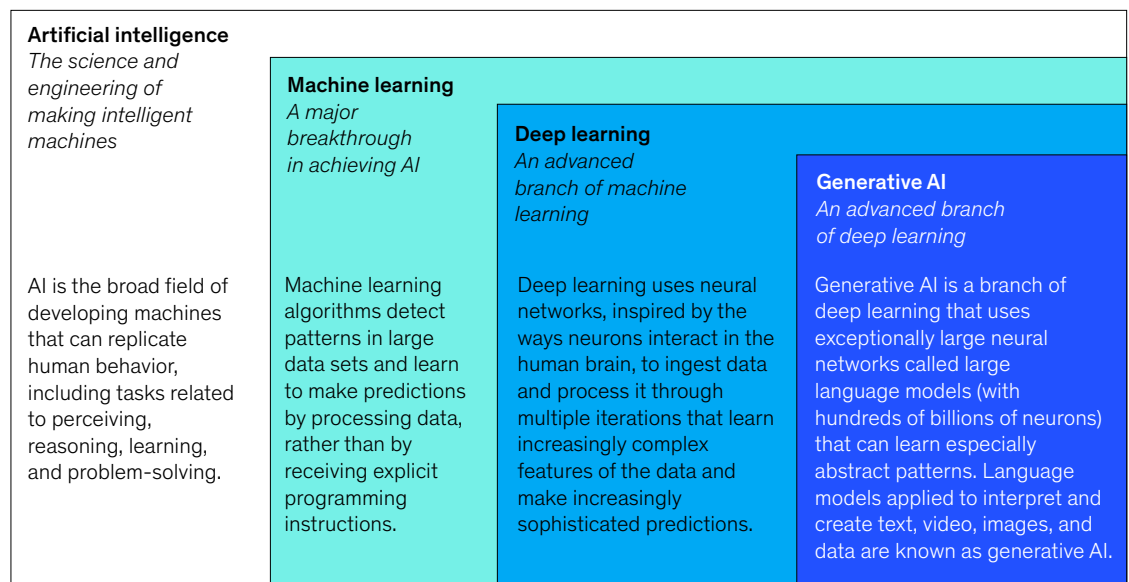
Deep learning is a more advanced version of machine learning that is particularly adept at processing a wider range of data resources (text as well as unstructured data including images), requires even less human intervention, and can often produce more accurate results than traditional machine learning. Deep learning uses neural networks—based on the [ways neurons interact in the human brain](#)—to ingest data and process it through multiple neuron layers that recognize increasingly complex features of the data. For example, an early layer might recognize something as being in a specific shape; building on this knowledge, a later layer might be able to identify the shape as a stop sign. Similar to machine learning, deep learning uses iteration to self-correct and improve its prediction capabilities. For example, once it “learns” what a stop sign looks like, it can recognize a stop sign in a new image.

Learn more about [QuantumBlack, AI by McKinsey](#).

Exhibit 1

Artificial intelligence is a machine’s ability to perform some cognitive functions we usually associate with human minds.

The evolution of artificial intelligence



The volume and complexity of data that is now being generated, too vast for humans to process and apply efficiently, has increased the potential of machine learning, as well as the need for it.

What is generative AI?

Generative AI (gen AI) is an AI model that generates content in response to a prompt. 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. Much is still unknown about gen AI's potential, but there are some questions we can answer—like how gen AI models are built, what kinds of problems they are best suited to solve, and how they fit into the broader category of AI and machine learning.

For more on generative AI and how it stands to affect business and society, check out our *Explainer* "[What is generative AI?](#)"

What is the history of AI?

The term "artificial intelligence" was [coined in 1956](#) by computer scientist John McCarthy for a workshop at Dartmouth. But he wasn't the first to write about the concepts we now describe as AI. Alan Turing introduced the concept of the "[imitation game](#)" in a 1950 paper. That's the test of a machine's ability to exhibit intelligent behavior, now known as the "Turing test." He believed researchers should focus on areas that don't require too much sensing and action, things like games and language translation. Research communities dedicated to concepts like computer vision, natural language understanding, and neural networks are, in many cases, several decades old.

MIT physicist Rodney Brooks [shared](#) details on the four previous stages of AI:

- **Symbolic AI (1956)**. Symbolic AI is also known as classical AI, or even GOF AI (good old-fashioned AI). The key concept here is the use of symbols and logical reasoning to solve problems. For example, we know [a German shepherd is a dog](#), which is a mammal; all mammals are warm-blooded; therefore, a German shepherd should be warm-blooded.

The main problem with symbolic AI is that humans still need to manually encode their knowledge of the world into the symbolic AI system, rather than allowing it to observe and encode relationships on its own. As a result, symbolic AI systems struggle with situations involving real-world complexity. They also lack the ability to learn from large amounts of data.

Symbolic AI was the dominant paradigm of AI research until the late 1980s.

- **Neural networks (1954, 1969, 1986, 2012)**. Neural networks are the technology behind the recent explosive growth of gen AI. Loosely modeling the [ways neurons interact in the human brain](#), neural networks ingest data and process it through multiple iterations that learn increasingly complex features of the data. The neural network can then make determinations

about the data, learn whether a determination is correct, and use what it has learned to make determinations about new data. For example, once it “learns” what an object looks like, it can recognize the object in a new image.

Neural networks were first proposed in 1943 in an academic paper by neurophysiologist Warren McCulloch and logician Walter Pitts. Decades later, in 1969, two MIT researchers mathematically demonstrated that neural networks could perform only very basic tasks. In 1986, there was another reversal, when computer scientist and cognitive psychologist Geoffrey Hinton and colleagues solved the neural network problem presented by the MIT researchers. In the 1990s, computer scientist Yann LeCun made major advancements in neural networks’ use in computer vision, while Jürgen Schmidhuber advanced the application of recurrent neural networks as used in language processing.

In 2012, Hinton and two of his students highlighted the power of deep learning. They applied Hinton’s algorithm to neural networks with many more layers than was typical, sparking a new focus on deep neural networks. These have been the main AI approaches of recent years.

- **Traditional robotics (1968).** During the first few decades of AI, researchers built robots to advance research. Some robots were mobile, moving around on wheels, while others were fixed, with articulated arms. Robots used the earliest attempts at computer vision to identify and navigate through their environments or to understand the geometry of objects and maneuver them. This could include moving around blocks of various shapes and colors. Most of these robots, just like the ones that have been used in factories for decades, rely on highly controlled environments with thoroughly scripted behaviors that they perform repeatedly.

Case study: Vistra and the Martin Lake Power Plant

Vistra is a large power producer in the United States, operating plants in 12 states with a capacity to power nearly 20 million homes. Vistra has committed to achieving net-zero emissions by 2050. In support of this goal, as well as to improve overall efficiency, [QuantumBlack, AI by McKinsey](#) worked with Vistra to build and deploy an AI-powered heat rate optimizer (HRO) at one of its plants.

“Heat rate” is a measure of the thermal efficiency of the plant; in other words, it’s the amount of fuel required to produce each unit of electricity. To reach the optimal heat rate, plant operators continuously monitor and tune hundreds of variables, such as steam temperatures, pressures, oxygen levels, and fan speeds.

Vistra and a McKinsey team, including data scientists and machine learning engineers, built a multilayered neural network model. The model combed through two years’ worth of data at the plant and learned which combination of factors would attain the most efficient heat rate at any point in time. When the models were accurate to 99 percent or higher and run through a rigorous set of real-world tests, the team converted them into an AI-powered engine that generates recommendations every 30 minutes for operators to improve the plant’s heat rate efficiency. One seasoned operations manager at the company’s plant in Odessa, Texas, said, “There are things that took me 20 years to learn about these power plants. This model learned them in an afternoon.”

Overall, the AI-powered HRO helped Vistra achieve the following:

- approximately 1.6 million metric tons of carbon abated annually
- 67 power generators optimized
- \$60 million saved in about a year

Read more about the Vistra story [here](#).

They have not contributed significantly to the advancement of AI itself.

But traditional robotics did have significant impact in one area, through a process called “simultaneous localization and mapping” (SLAM). SLAM algorithms helped contribute to self-driving cars and are used in consumer products like vacuum cleaning robots and quadcopter drones. Today, this work has evolved into behavior-based robotics, also referred to as haptic technology because it responds to human touch.

- **Behavior-based robotics (1985).** In the real world, there aren’t always clear instructions for navigation, decision making, or problem-solving. Insects, researchers observed, navigate very well (and are evolutionarily very successful) with few neurons. Behavior-based robotics researchers took inspiration from this, looking for ways robots could solve problems with partial knowledge and conflicting instructions. These behavior-based robots are embedded with neural networks.

Learn more about [QuantumBlack, AI by McKinsey](#).

What is artificial general intelligence?

The term “artificial general intelligence” (AGI) was coined to describe AI systems that possess [capabilities comparable to those of a human](#). In theory, AGI could someday replicate human-like cognitive abilities including reasoning, problem-solving, perception, learning, and language comprehension. But let’s not get ahead of ourselves: the key word here is “someday.” Most researchers and academics believe we are decades away from realizing AGI; some even predict we won’t see AGI this century, or ever. Rodney Brooks, an MIT roboticist and cofounder of iRobot, doesn’t believe AGI will arrive until [the year 2300](#).

The timing of AGI’s emergence may be uncertain. But when it does emerge—and it likely will—it’s going to be a very big deal, in every aspect of

our lives. Executives should begin working to understand the path to machines achieving human-level intelligence now and making the transition to a more automated world.

For more on AGI, including the four previous attempts at AGI, read our [Explainer](#).

What is narrow AI?

Narrow AI is the application of AI techniques to a specific and well-defined problem, such as chatbots like ChatGPT, algorithms that spot fraud in credit card transactions, and natural-language-processing engines that quickly process thousands of legal documents. Most current AI applications fall into the category of narrow AI. AGI is, by contrast, AI that’s intelligent enough to perform a broad range of tasks.

Learn more about [QuantumBlack, AI by McKinsey](#).

How is the use of AI expanding?

AI is a big story for all kinds of businesses, but some companies are clearly moving [ahead of the pack](#). Our state of AI in 2022 survey showed that adoption of AI models has more than doubled since 2017—and investment has increased apace. What’s more, the specific areas in which companies see value from AI have evolved, from manufacturing and risk to the following:

- marketing and sales
- product and service development
- strategy and corporate finance

One group of companies is pulling ahead of its competitors. Leaders of these organizations consistently make larger investments in AI, level up their practices to scale faster, and hire and upskill the best AI talent. More specifically, they link AI strategy to business outcomes and “[industrialize](#)” AI operations by designing modular data architecture that can quickly accommodate new applications.

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

We have yet to see the longtail effect of gen AI models. This means there are some inherent risks involved in using them—both known and unknown.

The outputs gen 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 other biases of the internet and society more generally).

It can also be manipulated to enable unethical or criminal activity. Since gen AI models burst onto the scene, organizations have become aware of users trying to “jailbreak” the models—that means trying to get them to break their own rules and deliver biased, harmful, misleading, or even illegal content. Gen AI organizations are responding to this threat in two ways: for one thing, they're collecting feedback from users on inappropriate content. They're also combing through their databases, identifying prompts that led to inappropriate content, and training the model against these types of generations.

But awareness and even action don't guarantee that harmful content won't slip the dragnet. Organizations that rely on gen AI models should be aware of the reputational and legal risks involved in unintentionally publishing biased, offensive, or copyrighted content.

These risks can be mitigated, however, in a few ways. “Whenever you use a model,” says McKinsey partner Marie El Hoyek, “you need to be able to [counter biases](#) and instruct it not to use inappropriate or flawed sources, or things you don't trust.” How? For one thing, 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 gen 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.

It's also important to keep a human in the loop (that is, to make sure a real human checks the output of a gen AI model before it is published or used) and avoid using gen 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 continue to change rapidly in the coming years. As gen AI becomes increasingly incorporated into business, society, and our personal lives, we can also expect a new regulatory climate to take shape. As organizations experiment—and create value—with these tools, leaders will do well to keep a finger on the pulse of regulation and risk.

Learn more about [QuantumBlack, AI by McKinsey](#).

What is the AI Bill of Rights?

The Blueprint for an AI Bill of Rights, prepared by the US government in 2022, provides a framework for how government, technology companies, and citizens can collectively ensure more accountable AI. As AI has become more ubiquitous, [concerns have surfaced](#) about a potential lack of transparency surrounding the functioning of gen AI systems, the data used to train them, issues of bias and fairness, potential intellectual property infringements, privacy violations, and more. The Blueprint comprises five principles that [the White House says](#) should “guide the design, use, and deployment of automated systems to protect [users] in the age of artificial intelligence.” They are as follows:

- ***The right to safe and effective systems.*** Systems should undergo predeployment testing, risk identification and mitigation, and ongoing monitoring to demonstrate that they are adhering to their intended use.
- ***Protections against discrimination by algorithms.*** Algorithmic discrimination is when automated systems contribute to unjustified different treatment of people based on their race, color, ethnicity, sex, religion, age, and more.

- *Protections against abusive data practices*, via built-in safeguards. Users should also have agency over how their data is used.
- *The right to know that an automated system is being used*, and a clear explanation of how and why it contributes to outcomes that affect the user.
- *The right to opt out*, and access to a human who can quickly consider and fix problems.

At present, more than 60 countries or blocs have national strategies governing the responsible use of AI (Exhibit 2). These include Brazil, China, the European Union, Singapore, South Korea, and the United States. The approaches taken vary from guidelines-based approaches, such as the Blueprint for an AI Bill of Rights in the United States, to comprehensive AI regulations that align with existing data protection and cybersecurity regulations, such as the EU's AI Act, due in 2024.

There are also collaborative efforts between countries to set out standards for AI use. The US–EU Trade and Technology Council is working toward greater alignment between Europe and the United States. The Global Partnership on Artificial Intelligence, formed in 2020, has 29 members including Brazil, Canada, Japan, the United States, and several European countries.

Even though AI regulations are still being developed, organizations should act now to avoid legal, reputational, organizational, and financial risks. In an environment of public concern, a misstep could be costly. Here are four no-regrets, preemptive actions organizations can implement today:

- **Transparency.** Create an inventory of models, classifying them in accordance with regulation, and record all usage across the organization that is clear to those inside and outside the organization.
- **Governance.** Implement a governance structure for AI and gen AI that ensures sufficient oversight, authority, and accountability both within the organization and with third parties and regulators.
- **Data, model, and technology management.**
 - **Data management.** Proper data management includes awareness of data sources, data classification, data quality and lineage, intellectual property, and privacy management.
 - **Model management.** Organizations should establish principles and guardrails for AI development and use them to ensure all AI models uphold fairness and bias controls.

Exhibit 2

Regulations related to AI governance vary around the world.

As of November 2023, nonexhaustive

Type of policy: Nonbinding principles (eg, OECD)

- Japan
- Singapore
- United Arab Emirates
- United Kingdom
- United States

General AI legislation proposed or being finalized

- Brazil
- Canada
- China
- South Korea
- European Union

Example countries without general AI legislation

- Australia
- India
- New Zealand
- Saudi Arabia

Source: OECD; McKinsey analysis

- **Cybersecurity and technology management.** Establish strong cybersecurity and technology to ensure a secure environment where unauthorized access or misuse is prevented.

- **Individual rights.** Make users aware when they are interacting with an AI system, and provide clear instructions for use.

How can organizations scale up their AI efforts from ad hoc projects to full integration?

Most organizations are dipping a toe into the AI pool—not cannonballing. Slow progress toward widespread adoption is likely due to cultural and organizational barriers. But leaders who effectively break down these barriers will be best placed to capture the opportunities of the AI era. And—crucially—companies that can’t take full advantage of AI are already being **sidelined** by those that can, in industries like auto manufacturing and financial services.

To scale up AI, organizations can make **three major shifts**:

1. **Move from siloed work to interdisciplinary collaboration.** AI projects shouldn’t be limited to discrete pockets of organizations. Rather, AI has the biggest impact when it’s employed by cross-functional teams with a mix of skills and perspectives, enabling AI to address broad business priorities.
2. **Empower frontline data-based decision making.** AI has the potential to enable faster, better decisions at all levels of an organization. But for this to work, people at all levels need to trust the algorithms’ suggestions and feel empowered to make decisions. (Equally, people should be able to override the algorithm or make suggestions for improvement when necessary.)
3. **Adopt and bolster an agile mindset.** The agile test-and-learn mindset will help reframe mistakes as sources of discovery, allaying the fear of failure and speeding up development.

Learn more about [QuantumBlack, AI](#) by McKinsey, and check out [AI-related job opportunities](#) if you’re interested in working at McKinsey.

Articles referenced:

- [“As gen AI advances, regulators—and risk functions—rush to keep pace,”](#) December 21, 2023, Andreas Kremer, [Angela Luget](#), [Daniel Mikkelsen](#), [Henning Soller](#), Malin Strandell-Jansson, and Sheila Zingg
- [“What is generative AI?,”](#) January 19, 2023
- [“Tech highlights from 2022—in eight charts,”](#) December 22, 2022
- [“Generative AI is here: How tools like ChatGPT could change your business,”](#) December 20, 2022, [Michael Chui](#), [Roger Roberts](#), and [Lareina Yee](#)
- [“The state of AI in 2022—and a half decade in review,”](#) December 6, 2022, [Michael Chui](#), [Bryce Hall](#), [Helen Mayhew](#), [Alex Singla](#), and [Alex Sukharevsky](#)
- [“Why businesses need explainable AI—and how to deliver it,”](#) September 29, 2022, [Liz Grennan](#), [Andreas Kremer](#), [Alex Singla](#), and [Peter Zipparo](#)
- [“Why digital trust truly matters,”](#) September 12, 2022, [Jim Boehm](#), [Liz Grennan](#), [Alex Singla](#), and [Kate Smaje](#)
- [“McKinsey Technology Trends Outlook 2023,”](#) July 20, 2023, [Michael Chui](#), [Mena Issler](#), [Roger Roberts](#), and [Lareina Yee](#)
- [“An AI power play: Fueling the next wave of innovation in the energy sector,”](#) May 12, 2022, [Barry Boswell](#), [Sean Buckley](#), [Ben Elliott](#), [Matias Melero](#), and [Micah Smith](#)
- [“Scaling AI like a tech native: The CEO’s role,”](#) October 13, 2021, [Jacomo Corbo](#), [David Harvey](#), [Nicolas Hohn](#), [Kia Javanmardian](#), and [Nayur Khan](#)

Find more content like this on the
McKinsey Insights App



Scan • Download • Personalize



- “[What the draft European Union AI regulations mean for business](#),” August 10, 2021, Misha Benjamin, [Kevin Buehler](#), Rachel Dooley, and Peter Zipparo
- “[Winning with AI is a state of mind](#),” April 30, 2021, [Thomas Meakin](#), Jeremy Palmer, [Valentina Sartori](#), and Jamie Vickers
- “[Breaking through data-architecture gridlock to scale AI](#),” January 26, 2021, [Sven Blumberg](#), [Jorge Machado](#), [Henning Soller](#), and [Asin Tavakoli](#)
- “[An executive’s guide to AI](#),” November 17, 2020, [Michael Chui](#), Brian McCarthy, and Vishnu Kamalnath
- “[Executive’s guide to developing AI at scale](#),” October 28, 2020, [Nayur Khan](#), Brian McCarthy, and Adi Pradhan
- “[An executive primer on artificial general intelligence](#),” April 29, 2020, [Federico Berruti](#), Pieter Nel, and Rob Whiteman
- “[The analytics academy: Bridging the gap between human and artificial intelligence](#),” *McKinsey Quarterly*, September 25, 2019, Solly Brown, Darshit Gandhi, [Louise Herring](#), and [Ankur Puri](#)

This article was updated in April 2024; it was originally published in April 2023.

Get to know and directly engage with senior McKinsey experts on AI

[Sven Blumberg](#) is a senior partner in McKinsey’s Düsseldorf office; [Michael Chui](#) is a partner at the McKinsey Global Institute and is based in the Bay Area office, where [Lareina Yee](#) is a senior partner; [Kia Javanmardian](#) is a senior partner in the Chicago office, where [Alex Singla](#), the global leader of QuantumBlack, AI by McKinsey, is also a senior partner; [Kate Smaje](#) and [Alex Sukharevsky](#) are senior partners in the London office.

Designed by McKinsey Global Publishing
Copyright © 2024 McKinsey & Company. All rights reserved.