Artificial Intelligence

Artificial intelligence (AI) is technology that enables computers and machines to simulate human learning, comprehension, problem solving, decision-making, creativity and autonomy.

By IBM

Jan 24, 2025 06:26 PM · 13 min. read · View original

What is AI?

Artificial intelligence (AI) is technology that enables computers and machines to simulate human learning, comprehension, problem solving, decision making, creativity and autonomy.

Applications and devices equipped with AI can see and identify objects. They can understand and respond to human language. They can learn from new information and experience. They can make detailed recommendations to users and experts. They can act independently, replacing the need for human

intelligence or intervention (a classic example being a self-driving car).

But in 2024, most AI researchers and practitioners—and most AI-related headlines—are focused on breakthroughs in generative AI (gen AI), a technology that can create original text, images, video and other content. To fully understand generative AI, it's important to first understand the technologies on which generative AI tools are built: machine learning (ML) and deep learning.

Machine learning

A simple way to think about AI is as a series of nested or derivative concepts that have emerged over more than 70 years:

Directly underneath AI, we have machine learning, which involves creating models by training an algorithm to make predictions or decisions based on data. It encompasses a broad range of techniques that

enable computers to learn from and make inferences based on data without being explicitly programmed for specific tasks.

There are many types of machine learning techniques or algorithms, including linear regression, logistic regression, decision trees, random forest, support vector machines (SVMs), k-nearest neighbor (KNN), clustering and more. Each of these approaches is suited to different kinds of problems and data.

But one of the most popular types of machine learning algorithm is called a neural network (or artificial neural network). Neural networks are modeled after the human brain's structure and function. A neural network consists of interconnected layers of nodes (analogous to neurons) that work together to process and analyze complex data. Neural networks are well suited to tasks that involve identifying complex patterns and relationships in large amounts of data.

The simplest form of machine learning is called supervised learning, which involves the use of labeled data sets to train algorithms to classify data or predict outcomes accurately. In supervised learning, humans pair each training example with an output label. The goal is for the model to learn the mapping between inputs and outputs in the training data, so it can predict the labels of new, unseen data.

Deep learning

Deep learning is a subset of machine learning that uses multilayered neural networks, called deep neural networks, that more closely simulate the complex decision-making power of the human brain.

Deep neural networks include an input layer, at least three but usually hundreds of hidden layers, and an output layer, unlike neural networks used in classic machine learning models, which usually have only one or two hidden layers.

These multiple layers enable unsupervised learning: they can automate the extraction of features from large, unlabeled and unstructured data sets, and make their own predictions about what the data represents.

Because deep learning doesn't require human intervention, it enables machine learning at a tremendous scale. It is well suited to natural language processing (NLP), computer vision, and other tasks that involve the fast, accurate identification complex patterns and relationships in large amounts of data. Some form of deep learning powers most of the artificial intelligence (AI) applications in our lives today.

Deep learning also enables:

- Semi-supervised learning, which combines supervised and unsupervised learning by using both labeled and unlabeled data to train AI models for classification and regression tasks.

- Self-supervised learning, which generates implicit labels from unstructured data, rather than relying on labeled data sets for supervisory signals.

- Reinforcement learning, which learns by trial-and-error and reward functions rather than by extracting information from hidden patterns.

- Transfer learning, in which knowledge gained through one task or data set is used to improve model performance on another related task or different data set.

Generative AI

Generative AI, sometimes called "gen AI", refers to deep learning models that can create complex original content—such as long-form text, high-quality images, realistic video or audio and more—in response to a user's prompt or request.

At a high level, generative models encode a simplified representation of their training data, and then draw from that representation to create new work that's similar, but not identical, to the original data.

Generative models have been used for years in statistics to analyze numerical data. But over the last decade, they evolved to analyze and generate more

complex data types. This evolution coincided with the emergence of three sophisticated deep learning model types:

- Variational autoencoders or VAEs, which were introduced in 2013, and enabled models that could generate multiple variations of content in response to a prompt or instruction.

- Diffusion models, first seen in 2014, which add "noise" to images until they are unrecognizable, and then remove the noise to generate original images in response to prompts.

- Transformers (also called transformer models), which are trained on sequenced data to generate extended sequences of content (such as words in sentences, shapes in an image, frames of a video or commands in software code). Transformers are at the core of most of today's headline-making generative AI tools, including ChatGPT and GPT-4, Copilot, BERT, Bard and Midjourney.

How generative AI works

In general, generative AI operates in three phases:

1. **Training,** to create a foundation model.
2. **Tuning,** to adapt the model to a specific application.
3. **Generation, evaluation and more tuning,** to improve accuracy.

Training

Generative AI begins with a "foundation model"; a deep learning model that serves as the basis for multiple different types of generative AI applications.

The most common foundation models today are large language models (LLMs), created for text generation applications. But there are also foundation models for image, video, sound or music generation, and multimodal foundation models that support several kinds of content.

To create a foundation model, practitioners train a deep learning algorithm on huge volumes of relevant raw, unstructured, unlabeled data, such as terabytes or petabytes of data text or images or video from the internet. The training yields a neural network of billions of *parameters*—encoded representations of the entities, patterns and relationships in the data—that can generate content autonomously in response to prompts. This is the foundation model.

This training process is compute-intensive, time-consuming and expensive. It requires thousands of clustered graphics processing units (GPUs) and weeks of processing, all of which typically costs millions of dollars. Open source foundation model projects, such as Meta's Llama-2, enable gen AI developers to avoid this step and its costs.

Tuning

Next, the model must be tuned to a specific content generation task. This can be done in various ways, including:

- Fine-tuning, which involves feeding the model application-specific labeled data—questions or prompts the application is likely to receive, and corresponding correct answers in the wanted format.

- Reinforcement learning with human feedback (RLHF), in which human users evaluate the accuracy or relevance of model outputs so that the model can improve itself. This can be as simple as having people type or talk back corrections to a chatbot or virtual assistant.

Generation, evaluation and more tuning

Developers and users regularly assess the outputs of their generative AI apps, and further tune the model—even as often as once a week—for greater accuracy or relevance. In contrast, the foundation model itself is updated much less frequently, perhaps every year or 18 months.

Another option for improving a gen AI app's performance is retrieval augmented generation (RAG), a technique for extending the foundation model to use relevant sources outside of the training data to refine the parameters for greater accuracy or relevance.

Benefits of AI

AI offers numerous benefits across various industries and applications. Some of the most commonly cited benefits include:

- **Automation of repetitive tasks.**
- **More and faster insight from data.**
- **Enhanced decision-making.**
- **Fewer human errors.**
- **24x7 availability.**
- **Reduced physical risks.**

Automation of repetitive tasks

AI can automate routine, repetitive and often tedious tasks—including digital tasks such as data collection, entering and preprocessing, and physical tasks such as warehouse stock-picking and manufacturing processes. This automation frees to work on higher value, more creative work.

Enhanced decision-making

Whether used for decision support or for fully automated decision-making, AI enables faster, more accurate predictions and reliable, data-driven decisions. Combined with automation, AI enables businesses to act on opportunities and respond to crises as they emerge, in real time and without human intervention.

Fewer human errors

AI can reduce human errors in various ways, from guiding people through the proper steps of a process, to flagging potential errors before they occur, and fully automating processes without human intervention. This is especially important in industries such as

healthcare where, for example, AI-guided surgical robotics enable consistent precision.

Machine learning algorithms can continually improve their accuracy and further reduce errors as they're exposed to more data and "learn" from experience.

Round-the-clock availability and consistency

AI is always on, available around the clock, and delivers consistent performance every time. Tools such as AI chatbots or virtual assistants can lighten staffing demands for customer service or support. In other applications—such as materials processing or production lines—AI can help maintain consistent work quality and output levels when used to complete repetitive or tedious tasks.

Reduced physical risk

By automating dangerous work—such as animal control, handling explosives, performing tasks in deep ocean water, high altitudes or in outer space—AI can eliminate the need to put human workers at risk of injury or worse. While they have yet to be perfected, self-driving cars and other vehicles offer the potential to reduce the risk of injury to passengers.

AI use cases

The real-world applications of AI are many. Here is just a small sampling of use cases across various

industries to illustrate its potential:

Customer experience, service and support

Companies can implement AI-powered chatbots and virtual assistants to handle customer inquiries, support tickets and more. These tools use natural language processing (NLP) and generative AI capabilities to understand and respond to customer questions about order status, product details and return policies.

Chatbots and virtual assistants enable always-on support, provide faster answers to frequently asked questions (FAQs), free human agents to focus on higher-level tasks, and give customers faster, more consistent service.

Fraud detection

Machine learning and deep learning algorithms can analyze transaction patterns and flag anomalies, such as unusual spending or login locations, that indicate fraudulent transactions. This enables organizations to respond more quickly to potential fraud and limit its impact, giving themselves and customers greater peace of mind.

Personalized marketing

Retailers, banks and other customer-facing companies can use AI to create personalized customer experiences and marketing campaigns that delight

customers, improve sales and prevent churn. Based on data from customer purchase history and behaviors, deep learning algorithms can recommend products and services customers are likely to want, and even generate personalized copy and special offers for individual customers in real time.

Human resources and recruitment

AI-driven recruitment platforms can streamline hiring by screening resumes, matching candidates with job descriptions, and even conducting preliminary interviews using video analysis. These and other tools can dramatically reduce the mountain of administrative paperwork associated with fielding a large volume of candidates. It can also reduce response times and time-to-hire, improving the experience for candidates whether they get the job or not.

Application development and modernization

Generative AI code generation tools and automation tools can streamline repetitive coding tasks associated with application development, and accelerate the migration and modernization (reformatting and replatorming) of legacy applications at scale. These tools can speed up tasks, help ensure code consistency and reduce errors.

Predictive maintenance

Machine learning models can analyze data from sensors, Internet of Things (IoT) devices and operational technology (OT) to forecast when maintenance will be required and predict equipment failures before they occur. AI-powered preventive maintenance helps prevent downtime and enables you to stay ahead of supply chain issues before they affect the bottom line.

AI challenges and risks

Organizations are scrambling to take advantage of the latest AI technologies and capitalize on AI's many benefits. This rapid adoption is necessary, but adopting and maintaining AI workflows comes with challenges and risks.

Data risks

AI systems rely on data sets that might be vulnerable to data poisoning, data tampering, data bias or cyberattacks that can lead to data breaches. Organizations can mitigate these risks by protecting data integrity and implementing security and availability throughout the entire AI lifecycle, from development to training and deployment and postdeployment.

Model risks

Threat actors can target AI models for theft, reverse engineering or unauthorized manipulation. Attackers

might compromise a model's integrity by tampering with its architecture, weights or parameters; the core components that determine a model's behavior, accuracy and performance.

Operational risks

Like all technologies, models are susceptible to operational risks such as model drift, bias and breakdowns in the governance structure. Left unaddressed, these risks can lead to system failures and cybersecurity vulnerabilities that threat actors can use.

Ethics and legal risks

If organizations don't prioritize safety and ethics when developing and deploying AI systems, they risk committing privacy violations and producing biased outcomes. For example, biased training data used for hiring decisions might reinforce gender or racial stereotypes and create AI models that favor certain demographic groups over others.

AI ethics and governance

AI ethics is a multidisciplinary field that studies how to optimize AI's beneficial impact while reducing risks and adverse outcomes. Principles of AI ethics are applied through a system of AI governance consisted of guardrails that help ensure that AI tools and systems remain safe and ethical.

AI governance encompasses oversight mechanisms that address risks. An ethical approach to AI governance requires the involvement of a wide range of stakeholders, including developers, users, policymakers and ethicists, helping to ensure that AI-related systems are developed and used to align with society's values.

Here are common values associated with AI ethics and responsible AI:

Explainability and interpretability

As AI becomes more advanced, humans are challenged to comprehend and retrace how the algorithm came to a result. Explainable AI is a set of processes and methods that enables human users to interpret, comprehend and trust the results and output created by algorithms.

Fairness and inclusion

Although machine learning, by its very nature, is a form of statistical discrimination, the discrimination becomes objectionable when it places privileged groups at systematic advantage and certain unprivileged groups at systematic disadvantage, potentially causing varied harms. To encourage fairness, practitioners can try to minimize algorithmic bias across data collection and model design, and to build more diverse and inclusive teams.

Robustness and security

Robust AI effectively handles exceptional conditions, such as abnormalities in input or malicious attacks, without causing unintentional harm. It is also built to withstand intentional and unintentional interference by protecting against exposed vulnerabilities.

Accountability and transparency

Organizations should implement clear responsibilities and governance structures for the development, deployment and outcomes of AI systems. In addition, users should be able to see how an AI service works, evaluate its functionality, and comprehend its strengths and limitations. Increased transparency provides information for AI consumers to better understand how the AI model or service was created.

Privacy and compliance

Many regulatory frameworks, including GDPR, mandate that organizations abide by certain privacy principles when processing personal information. It is crucial to be able to protect AI models that might contain personal information, control what data goes into the model in the first place, and to build adaptable systems that can adjust to changes in regulation and attitudes around AI ethics.

Weak AI vs. Strong AI

In order to contextualize the use of AI at various levels of complexity and sophistication, researchers have defined several types of AI that refer to its level of sophistication:

Weak AI: Also known as “narrow AI,” defines AI systems designed to perform a specific task or a set of tasks. Examples might include “smart” voice assistant apps, such as Amazon’s Alexa, Apple’s Siri, a social media chatbot or the autonomous vehicles promised by Tesla.

Strong AI: Also known as “artificial general intelligence” (AGI) or “general AI,” possess the ability to understand, learn and apply knowledge across a wide range of tasks at a level equal to or surpassing human intelligence. This level of AI is currently theoretical and no known AI systems approach this level of sophistication. Researchers argue that if AGI is even possible, it requires major increases in computing power. Despite recent advances in AI development, self-aware AI systems of science fiction remain firmly in that realm.

History of AI

The idea of "a machine that thinks" dates back to ancient Greece. But since the advent of electronic computing (and relative to some of the topics

discussed in this article) important events and milestones in the evolution of AI include the following:

1950

Alan Turing publishes Computing Machinery and Intelligence. In this paper, Turing—famous for breaking the German ENIGMA code during WWII and often referred to as the "father of computer science"—asks the following question: "Can machines think?"

From there, he offers a test, now famously known as the "Turing Test," where a human interrogator would try to distinguish between a computer and human text response. While this test has undergone much scrutiny since it was published, it remains an important part of the history of AI, and an ongoing concept within philosophy as it uses ideas around linguistics.

1956

John McCarthy coins the term "artificial intelligence" at the first-ever AI conference at Dartmouth College. (McCarthy went on to invent the Lisp language.) Later that year, Allen Newell, J.C. Shaw and Herbert Simon create the Logic Theorist, the first-ever running AI computer program.

1967

Frank Rosenblatt builds the Mark 1 Perceptron, the first computer based on a neural network that "learned" through trial and error. Just a year later,

Marvin Minsky and Seymour Papert publish a book titled Perceptrons, which becomes both the landmark work on neural networks and, at least for a while, an argument against future neural network research initiatives.

1980

Neural networks, which use a backpropagation algorithm to train itself, became widely used in AI applications.

1995

Stuart Russell and Peter Norvig publish Artificial Intelligence: A Modern Approach, which becomes one of the leading textbooks in the study of AI. In it, they delve into four potential goals or definitions of AI, which differentiates computer systems based on rationality and thinking versus acting.

1997

IBM's Deep Blue beats then world chess champion Garry Kasparov, in a chess match (and rematch).

2004

John McCarthy writes a paper, What Is Artificial Intelligence?, and proposes an often-cited definition of AI. By this time, the era of big data and cloud computing is underway, enabling organizations to manage ever-larger data estates, which will one day be used to train AI models.

2011

IBM Watson® beats champions Ken Jennings and Brad Rutter at Jeopardy! Also, around this time, data science begins to emerge as a popular discipline.

2015

Baidu's Minwa supercomputer uses a special deep neural network called a convolutional neural network to identify and categorize images with a higher rate of accuracy than the average human.

2016

DeepMind's AlphaGo program, powered by a deep neural network, beats Lee Sodol, the world champion Go player, in a five-game match. The victory is significant given the huge number of possible moves as the game progresses (over 14.5 trillion after just four moves). Later, Google purchased DeepMind for a reported USD 400 million.

2022

A rise in large language models or LLMs, such as OpenAI's ChatGPT, creates an enormous change in performance of AI and its potential to drive enterprise value. With these new generative AI practices, deep-learning models can be pretrained on large amounts of data.

2024

The latest AI trends point to a continuing AI

renaissance. Multimodal models that can take multiple types of data as input are providing richer, more robust experiences. These models bring together computer vision image recognition and NLP speech recognition capabilities. Smaller models are also making strides in an age of diminishing returns with massive models with large parameter counts.