Co-funded by the European Union

Erasmus+
Enriching lives, opening minds.

The History of Artificial Intelligence

From the first dreams of thinking machines to everyday AI — a plain-language guide

Introduction

Artificial intelligence, or AI, is built on the idea that computers can think and learn — much like people. Today AI helps us every day: it recommends films on Netflix, powers the voice assistants on our phones, and talks to us through online chatbots.

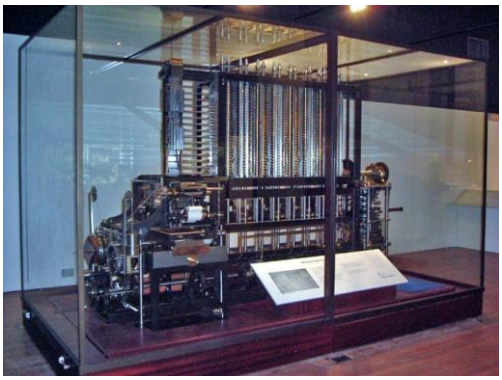
How did we get here? The story of AI is woven through with grand ambitions, long winters, surprising breakthroughs, and a future still being written. This article takes you through that journey in plain, accessible language.

Before the Beginning: Early Dreamers and Victorian Visionaries

Charles Babbage and the Mechanical Brain (1820s – 1871)

Long before the word 'computer' existed, the English mathematician Charles Babbage was already dreaming of thinking machines. In the 1820s he designed the Difference Engine — a mechanical calculator assembled from thousands of interlocking gears, intended to compute mathematical tables without human error. It was the world's first automatic calculating machine. Babbage never completed it in his lifetime, but in 1991 the London Science Museum built one from his original drawings, and it worked perfectly.

Even more ambitious was Babbage's Analytical Engine, designed from 1837 onwards. It was a truly revolutionary invention: a general-purpose machine that could be programmed using punched cards. It had a memory (called the 'store') and a processing unit (the 'mill'). It could perform any calculation, follow conditional logic, and loop through instructions — concepts that are today at the heart of every computer and AI system.



© Joe D, Wikimedia Commons, CC BY-SA 1.0

Babbage was also a genuine philosophical pioneer of AI. In his 1837 work 'Ninth Bridgewater Treatise' he explored whether a machine could simulate the laws of the universe, approaching the idea that thought itself might be mechanical. His collaborator Ada Lovelace — often called the world's first computer programmer — speculated that the Analytical Engine might become 'a machine that thinks and reasons', but cautioned that machines can only do what people program them to do. That debate continues to this day.

Funded by the European Union. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or Agency for Mobility and EU Programmes (AMEUP). Neither the European Union nor AMEUP can be held responsible for them.

Babbage's core insight — that logic can be encoded, that machines can reason, and that computing is the next great tool of humanity — makes him not only the father of the computer, but one of the earliest ancestors of AI.

A Forgotten Spanish AI Pioneer: Leonardo Torres Quevedo (1912)

Here is a name most of us have never heard — but absolutely should know: Leonardo Torres Quevedo. Working in his Automation Laboratory in Madrid in the early 1900s, this Spanish engineer from Cantabria did something that many historians now consider the world's first genuine act of artificial intelligence.

In 1912 he built El Ajedrecista — 'The Chess Player' — the world's first fully autonomous decision-making machine. Unlike earlier fake 'chess automata' (such as the famous Mechanical Turk of 1770, which concealed a real chess player inside), El Ajedrecista hid no human being: it was a genuine machine that played chess.

Using electromagnets, electrical sensors, and a clever algorithm, the machine could defeat any human opponent. It automatically detected the opponent's moves, calculated its own next move, and physically moved the pieces with a mechanical arm. If a player tried to cheat, a light would flash — and after three illegal moves, the machine would refuse to continue.

© MdeVicente, Wikimedia Commons, CC0

Torres Quevedo publicly demonstrated El Ajedrecista in Paris in 1914 to great sensation. In 1915, Scientific American published an article about his work: 'Torres and His Remarkable Automatic Devices. He Wishes to Substitute Machines for the Human Mind.' In a 1914 essay, 'Essays on Automatics', Torres went further, arguing that the boundary between what machines and humans can do needed to be fundamentally reconsidered. In that same essay he introduced floating-point arithmetic — a concept that would not be 'rediscovered' until decades later.

An improved version of El Ajedrecista was demonstrated in 1951 at the great Paris Computing Congress, where it defeated the chess grandmaster Saveliy Tartakower — becoming the first machine ever to beat a chess grandmaster. Norbert Wiener, the father of cybernetics, played against El Ajedrecista at the same event.

Both original machines survive in full working order and can be seen at the Torres Quevedo Museum at the Polytechnic University of Madrid. Sadly, English-speaking researchers largely overlooked Torres Quevedo's work, so he had little direct influence on the AI pioneers of the 1940s and 50s. History is slowly correcting that injustice.

Key events: Before modern AI

Year	Event	Description
1820s	Difference Engine	Charles Babbage designs the world's first automatic mechanical calculator.
1837	Analytical Engine	Babbage designs a general-purpose programmable machine — a forerunner of the modern computer.

1843	Ada Lovelace	Writes the first algorithm and speculates about machines that might 'think'.
1912	El Ajedrecista	Torres Quevedo builds the world's first autonomous decision-making machine — a chess automaton.
1914	Essays on Automatics	Torres Quevedo publishes a theory of thinking machines and introduces floating-point arithmetic.
1951	Grandmaster defeated	El Ajedrecista defeats chess grandmaster Tartakower in Paris — the first machine to achieve this.

Part 1: The Big Idea (1940s and 1950s)

How Did It All Begin?

The idea of thinking machines — of artificial beings — is ancient, reaching back thousands of years. The modern story of AI begins in the 1940s, when scientists started to understand how the human brain works and asked: could we build a machine that does the same?

In 1943, two researchers — Warren McCulloch and Walter Pitts — published a paper describing how brain cells (neurons) might work using simple mathematics. It was the first step towards building a computational brain.

In 1950, the British mathematician Alan Turing posed what became a famous question: 'Can machines think?' According to the Turing Test, if a machine can hold a conversation and you cannot tell with certainty whether you are talking to a human or a computer, then it is 'intelligent'. That simple idea shaped AI research for decades.

The Milestone: The Birth of AI (1956)

In the summer of 1956, a group of scientists held a workshop at Dartmouth College in the United States. One of them, John McCarthy, coined the term 'Artificial Intelligence' for the very first time. That gathering is now considered the official birth of AI as a field of research.

Part 2: Early Excitement — The First AI Era (1956–1974)

Optimism and Early Programmes

After Dartmouth, AI research advanced rapidly. Scientists were extraordinarily optimistic. They believed that within a single generation, machines would be able to do everything a human could.

Early AI programmes could solve mathematical problems, play draughts, and hold simple conversations. The most famous early chatbot, ELIZA, was created at MIT in 1966. ELIZA could imitate a therapist by turning whatever you said into a question. People were so impressed that some actually believed they were talking to a real person.

Another important programme was SHRDLU (1970), which could understand simple commands in English and move virtual objects around in a 'blocks world'. Researchers at the time believed they were only a few years away from fully intelligent machines.

Key events: Early AI

Year	Event	Description

1950	The Turing Test	Alan Turing proposes a test for measuring machine intelligence.
1956	The name 'AI'	John McCarthy coins the term 'Artificial Intelligence'.
1966	ELIZA chatbot	The first chatbot imitates human conversation.
1970	SHRDLU	A programme understands English commands in a simulated environment.

Part 3: The First AI Winter (1974–1980)

A Reality Check

By the mid-1970s, the first wave of excitement had hit a wall. Computers were slow and had very little memory. AI programmes that worked on simple problems failed completely when given anything more complex.

Funding dried up, and governments that had invested in AI research pulled back. This period of disappointment and reduced funding became known as an 'AI winter' — a time when progress almost froze.

The main problems were: computers that were not nearly powerful enough, a lack of data, and research methods that simply could not handle the complexity of the real world.

Part 4: Expert Systems — A New Hope (1980s)

Teaching Computers Rules

In the 1980s, AI came back to life through a new approach: expert systems. Instead of trying to teach computers to learn on their own, researchers encoded the knowledge of human experts into thousands of rules in the form of 'if this, then that'.

These systems could diagnose medical conditions, help configure computers, and offer advice on chemical structures. Companies began investing millions of dollars again, and Japan launched a massive national AI project — the Fifth Generation Computer Project — aimed at building the most advanced AI computers in the world.

For a while, things worked. Expert systems were genuinely useful for specific tasks. But they had a fatal flaw: they could only do what they had been explicitly programmed to do, and they could not learn anything new.

The Second AI Winter (Late 1980s and Early 1990s)

By the late 1980s, expert systems had proved too expensive to maintain and too narrow to expand. The specialised hardware they ran on became obsolete, and funding dried up once again. AI entered another cold period.

This time, however, researchers did not give up. They continued working quietly on new approaches — in particular, a technique called neural networks.

Part 5: Machine Learning Takes Over (1990s and 2000s)

Learning from Data

The shift from coding rules to teaching computers from examples — a technique called machine learning — triggered a quiet revolution in the 1990s.

Instead of writing thousands of rules for a spam filter, you could show a computer thousands of emails labelled 'spam' and 'not spam' and let it discover the patterns itself.

In 1997, IBM's chess computer Deep Blue defeated the reigning world chess champion Garry Kasparov. It was a decisive moment — proof that machines could beat humans at complex strategic games.

In 1998, researcher Yann LeCun and others developed convolutional neural networks (CNNs), which could recognise handwritten digits with high accuracy. It was an early sign of what neural networks could achieve.

Geoffrey Hinton — the British-Canadian computer scientist whom many regard as one of the 'godfathers of AI' — together with his research team at the University of Toronto and colleagues including Yann LeCun and Yoshua Bengio, dramatically revived the nearly forgotten neural networks in the mid-2000s, reshaping them into a powerful approach known as deep learning. This technique, which uses multi-layered neural networks capable of automatically learning complex patterns from vast amounts of data, launched a new era in AI development that profoundly transformed fields such as speech recognition, computer vision, and natural language processing.

Key events: The machine learning era

Year	Event	Description
1997	Deep Blue	IBM's computer defeats world chess champion Kasparov.
1998	Neural networks	LeCun develops CNNs for handwriting recognition.
2001	Web data	The internet explosion gives AI enormous new quantities of data to learn from.
2006	Deep learning	Geoffrey Hinton's team revives neural networks as 'deep learning'.

Part 6: The Deep Learning Revolution (2010–2020)

When Did AI Really Take Off?

The decade from 2010 to 2020 was truly revolutionary for AI. Three things came together at exactly the right moment: enormous amounts of data (thanks to the internet), far more powerful computers (especially graphics processing units, or GPUs), and improved algorithms for training neural networks.

In 2012, a neural network called AlexNet entered an image-recognition competition and crushed the competition — reducing the error rate by more than ten percentage points. The deep learning era had officially begun.

Suddenly, AI could recognise faces, understand speech, translate languages, and beat humans at games that had been considered too complex for computers.

Landmark Moments

In 2011, IBM's Watson won America's popular television quiz show Jeopardy!, defeating the greatest human champions. Watson had to understand complex, nuanced questions in English and answer them accurately. It did so brilliantly.

In 2016, DeepMind's AlphaGo defeated the world champion at the ancient board game Go. This was considered a monumental achievement — Go has more possible positions than there are atoms in the universe. Experts had thought this would take another decade; AlphaGo achieved it years ahead of schedule.

In 2014, Ian Goodfellow — an American AI researcher then studying for his doctorate in Montreal — invented Generative Adversarial Networks (GANs). The idea reportedly came to him at a party. The concept is based on a competition between two neural networks: a generator creates fake data, while a discriminator tries to tell the fake from the real — each constantly improving the other. This innovation opened up entirely new applications in art, medicine, and the entertainment industry.

In 2017, a Google research team published a paper entitled *"Attention Is All You Need"* introducing the Transformer architecture. Unlike earlier approaches, the Transformer introduced a self-attention mechanism that allows a model to consider all parts of a text simultaneously and recognise complex relationships between words. This architecture became the foundation of almost all modern language models — from GPT and BERT to Gemini and Claude — profoundly changing the field of artificial intelligence.

Virtual assistants became part of everyday life: Apple's Siri launched in 2011, Amazon's Alexa in 2014, and Google Assistant in 2016. AI had moved from research laboratories into living rooms.

Key events: The deep learning revolution

Year	Event	Description
2011	Watson wins	IBM's Watson defeats humans on TV quiz show Jeopardy!
2012	AlexNet	Deep learning shatters records in image-recognition competition.
2014	GANs	Ian Goodfellow invents Generative Adversarial Networks.
2016	AlphaGo	DeepMind's AI defeats the world champion at Go.
2017	Transformers	Google introduces the Transformer architecture — the foundation of modern AI.

Part 7: The Generative AI Era (2020 onwards)

AI That Creates

Perhaps the most dramatic turn in AI history came after 2020, when AI shifted from recognising things to creating them. This is called 'generative AI' — AI that can produce text, images, music, and code from scratch.

The key technology behind this is the Transformer, invented by Google researchers in 2017. It became the foundation of a new class of AI models called Large Language Models (LLMs) — systems trained on enormous quantities of text from the internet that can understand and generate human language with remarkable fluency.

ChatGPT and the Public Awakening

In November 2022, OpenAI released ChatGPT. Within five days it had one million users — faster than any product in history. Suddenly, everyone could talk to an AI, ask it to write an essay or code, explain complex topics, or generate ideas.

Google then released Gemini, Meta released Llama, and Anthropic released Claude, while dozens of other companies launched their own AI assistants. AI-generated images, videos, and music became ubiquitous.

AI Enters Every Industry

From 2024, AI has been spreading into almost every area of life. Doctors use it to detect cancer in medical scans, lawyers to review contracts, programmers to write software faster — while students, writers, designers, and scientists have found in AI a creative and intellectual partner.

Key events: The generative AI era

Year	Event	Description
2020	GPT-3	OpenAI's large language model demonstrates extraordinary text-generation capabilities.
2021	DALL-E	AI can generate realistic images from plain-language descriptions.
2022	ChatGPT	OpenAI's chatbot reaches one million users in five days.
2023	GPT-4 / Claude	More powerful AI models match or exceed humans on many tasks.
2024–25	AI everywhere	AI for coding, reasoning, and multimodal tasks comes into widespread use.

Part 8: Challenges and Open Questions

Problems We Need to Solve

The rapid development of AI has brought serious challenges that researchers, governments, and society are actively grappling with.

Bias and fairness: AI systems learn from data created by humans — which means they can absorb human prejudices. A hiring tool powered by AI might unfairly favour certain groups, while a facial-recognition system might work less well for some ethnic groups. Correcting these biases is a major challenge.

Misinformation: Generative AI can create convincing but false text, images, and video — so-called 'deepfakes'. This makes it harder to tell the real from the fake and erodes trust in media and public discourse.

Privacy: AI systems learn from vast quantities of data, much of it personal. Questions about who owns that data and how it should be used remain hotly debated.

Jobs: As AI automates more and more tasks, some jobs will change or disappear. History suggests that new technology always creates new jobs too — but the transition can be painful for many workers.

Safety: As AI systems grow more capable, aligning their behaviour with human values becomes increasingly important. This is an active area of research at companies including Anthropic, OpenAI, and DeepMind.

Conclusion: The Road Ahead

The history of AI is a story of grand ambitions, painful setbacks, and astonishing breakthroughs. What began with the question 'Can machines think?' has become one of the defining technological and social questions of our time.

AI has passed through multiple cycles of excitement and disappointment. Despite the fact that there is still an enormous amount of room to grow, we now live in what many call the most exciting period in the history of AI.

Machines will not soon replace human creativity, empathy, or judgement — but they are becoming powerful partners and tools that can extend human capability in ways we are only beginning to understand.

> The future of AI will be shaped not only by technology, but by the choices people make about how to build it, use it, and govern it. That future belongs to all of us.