Chapter 2
Artificial Intelligence: Definition and Background

2.1 Definitions of AI

If we want to embed AI in society, we need to understand what it is. What do we mean by artificial intelligence? How has the technology developed? Where do we stand now?

Defining AI is not easy; in fact, there is no generally accepted definition of the concept.[1] Numerous different ones are used, and this can easily lead to confusion. It is therefore important to clarify our use of the term. We start by discussing various definitions of AI, then explain which we have settled on. The sheer variety of definitions in circulation is not due to carelessness, but inherent in the phenomenon of AI itself.

In its broadest definition, AI is equated with algorithms. However, this is not an especially useful approach for our analysis. Algorithms predate AI and have been widely used outside this field. The term 'algorithm' is derived from the name of the ninth-century Persian mathematician Mohammed ibn Musa al-Kharizmi and refers to a specific instruction for solving a problem or performing a calculation. If we were to define AI simply as the use of algorithms, it would include many other activities such as the operations of a pocket calculator or even the instructions in a cookbook.

In its strictest definition, AI stands for the imitation by computers of the intelligence inherent in humans. Purists point out that many current applications are still relatively simple and therefore not true AI. That makes this definition inappropriate for our report, too; to use it would be to imply that AI does not exist at present. We would effectively be defining the phenomenon out of existence.

A common definition of AI is that it is a technology that enables machines to imitate various complex human skills. This, however, does not give is much to go on. In fact, it does no more than render the term 'artificial intelligence' in different

[1] Russell & Norvig, 2020.

H. Sheikh et al., *Mission AI*, Research for Policy,
https://doi.org/10.1007/978-3-031-21448-6_2

words. As long as those 'complex human skills' are not specified, it remains unclear exactly what AI is. The same applies to the definition of AI as the performance by computers of complex tasks in complex environments.

Other definitions go further in explaining these skills and tasks. For example, the computer scientist Nils John Nilsson describes a technology that "functions appropriately and with foresight in its environment".[2] Others speak of the ability to perceive, to pursue goals, to initiate actions and to learn from a feedback loop.[3] A similar definition has been put forward by the High-Level Expert Group on Artificial Intelligence (AI HLEG) of the European Commission (EC): "Systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals."[4]

These task-based definitions go some way towards giving us a better understanding of what AI is. But they still have limitations. Concepts like "some degree of autonomy" remain somewhat vague. Moreover, these definitions still seem overly broad in that they describe phenomena that most of us would not be inclined to bundle under the term AI. For example, Nilsson's definition also applies to a classic thermostat. This device is also able to perceive (measure the temperature of the room), pursue goals (the programmed temperature), initiate actions (regulate the thermostat) and learn from a feedback loop (stop once the programmed temperature has been reached). Even so, most people would not be inclined to regard a thermostat as AI.

It is not surprising that AI is so difficult to define clearly. It is, after all, an imitation or simulation of something we do not yet fully understand ourselves: human intelligence. This has long been the subject of research by psychologists, behavioural scientists and neurologists, amongst others. We know a lot about intelligence and the human brain, but that knowledge is far from complete and there is no consensus as to what exactly human intelligence is. Until that comes about, it is impossible to be precise about how that intelligence can be imitated artificially.

Moreover, there is a clear interface between research into human intelligence on the one hand and into artificial intelligence on the other, where our understanding of both is co-evolving. We can illustrate this using the example of chess, a game AI has been able to play extremely well since the 1990s. In the 1950s an expert predicted, "If one could devise a successful chess machine, one would seem to have penetrated to the core of human intellectual endeavour."[5] In 1965 the Russian mathematician Alexander Kronrod called chess "the fruit fly of intelligence" – that is, the key to understanding it.[6] So people were amazed when a computer did finally manage to beat a chess grandmaster. In the Netherlands, research in this field led to the

[2] Nilsson, 2009: 13.

[3] See, for example, DenkWerk, 2018.

[4] High-Level Expert Group on Artificial Intelligence, 2019. At the end of this document the authors expand on their initial definition with a detailed explanation of its various elements.

[5] Bostrom, 2016: 14.

[6] Floridi, 2014: 139.

founding of the Dutch Computer Chess Association foundation (Computer Schaak Vereniging Nederland, CSVN) in 1980. Amongst its initiators were chess legend and former world champion Max Euwe and computer scientist Jaap van den Herik. Three years later Van den Herik would defend the first PhD thesis in the Netherlands on computer chess and artificial intelligence. In 1997, when Garry Kasparov was defeated by Deep Blue, IBM's chess computer, the cover of *Newsweek* claimed that this was "The brain's last stand." Chess was considered the pinnacle of human intelligence. At first glance this is not surprising, because the game is difficult for people to learn and those who are good at it are often considered very clever. It was with this in mind that commentators declared Deep Blue's victory a huge breakthrough for human intelligence in machines, stating that it must now be within the reach of computers to surpass humans in all sorts of activities we consider easier than chess.

Yet this did not happen. We have since revised our view of this form of intelligence. Chess is not the crowning glory of human intellectual endeavour; it is simply a mathematical problem with very clear rules and a finite set of alternatives. In this sense, a chess program is actually not very different from a pocket calculator, which can also do things too difficult even for very clever people. But they do not make it an artificial form of human intelligence.

Chess was long considered an extremely advanced game. However, years of research have revealed that something as apparently simple as recognizing a cat in a photograph – which AI has only learnt to do in recent years – is far more complex. This phenomenon has come to be known as Moravec's paradox: certain things that are very difficult for humans, such as chess or advanced calculus, are quite easy for computers.[7] But things that are very simple for us humans, such as perceiving objects or using motor skills to do the washing up, turn out to be very difficult for computers: "It is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers [draughts], and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility."[8]

This reflects a recurring pattern in the history of AI: people's idea of what constitutes a complex form of human intelligence has evolved with the increasing skills of our computers. What used to be considered a fine example of artificial intelligence eventually degrades to a simple calculation that no longer deserves the name AI. Pamela McCorduck calls this the 'AI effect': as soon as a computer figures out how to do something, people declare that it is 'just a calculation' and not actual intelligence. According to Nick Bostrom, director of the Oxford Institute for Internet Governance, AI includes anything that impresses us at any given time. Once we are no longer impressed, we simply call it software.[9] A chess app on a smartphone is an

[7] Moravec, 1988. The AI scientist Donald Knuth formulated it differently. He noticed that AI could do things that humans need to think about but failed at tasks humans do without thinking, like recognizing objects, analysing images and moving an arm (Bostrom, 2016: 17).

[8] Moravec, 1988: 15.

[9] Bostrom, 2016.

example. The difficulties in defining AI are therefore not the result of some shortcoming or carelessness, but rather arise from the fact that we were long unable to determine precisely what intelligence we wanted to imitate artificially.

In this context, it is also claimed that the use of the term 'intelligence' is misleading in that it wrongly suggests that machines can do the same things as people. Some have therefore suggested adopting other terms. Agrawal, Gans and Goldfarb say that modern technology does not bring us intelligence, but only one of its components, predictions, and so they use the term 'prediction machines'.[10] The philosopher Daniel Dennett goes even further and suggests that we should not model AI on humans at all. These are not artificial people, but a completely new type of entity – one he compares with oracles: entities that make predictions, but unlike humans have no personality, conscience or emotions.[11] In other words, AI appears to do what people do but in fact does something else. Edsger Dijkstra illustrated this through the question 'Do submarines swim?'.[12] What these vessels do is similar to what humans call swimming, but to call it that would be a mistake. AI can certainly do things that look like the intelligent things we do, but in fact it does them very differently.

This perspective also sheds light on the Moravec paradox mentioned above. Recognizing faces is easy for humans, but difficult for computers. This is because recognizing others was critical for our evolutionary survival and so our brain has learned to do it without thinking.

Being able to play chess was not essential in evolution and is therefore more difficult to master. That is to say, it requires a certain level of computational skill. Computers have not evolved biologically, so their abilities are different from those of humans. One important aspect of this theory is that we should not try too hard to understand AI from the point of view of human intelligence. Nevertheless, the term 'artificial intelligence' has become so commonplace that there is no point trying to replace it now.

Finally, AI is also often equated with the latest technology. As we will see later, AI has gained huge momentum in recent years. One of the major drivers of this has been progress in a specific area of the field, 'machine learning' (ML), where the innovation has resulted in what is now called 'deep learning' (DL). It is this technology that has been behind recent milestones, such as computers able to recognize faces and play games like Go. By contrast with the more traditional approaches whereby computer systems apply fixed rules, ML and DL algorithms can recognize patterns in data. We also speak here of 'self-learning algorithms'. Many people who talk about AI today are actually referring to these algorithms, and often specifically

[10] Agrawal et al., 2018: 2, 39. Drawing on the work of Jeff Hawkins, these authors believe that the foundation of intelligence is 'prediction'.

[11] Dennett, 2019.

[12] Dignum, 2019.

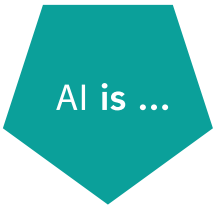
to DL. The focus on this technology is important because several pressing questions concerning AI are particularly relevant here (such as problems of explainability).

Given all the different definitions discussed here and elsewhere, we have settled on an open definition of AI. Two considerations are relevant in this respect. Firstly, it would be unwise for the purposes of this report to limit the definition of AI to a specific part of the technology. If, for example, we were to confine ourselves to 'deep learning' as discussed above, we would ignore the fact that many current issues also play a role in other AI domains, such as logical systems. One such example is the 'black box' question. Also, most applications of AI used by governments are not based on advanced techniques like DL and yet still have many important issues that need to be addressed in this report. Too narrow a definition would place them outside the scope of this study. While developments in DL have indeed resulted in a great leap forward, moreover, at the end of the next chapter we also point out several shortcomings of this technique. In fact, future advances in AI may well come from other fields. To allow for this, it is important to have an open definition of AI.

Secondly, as discussed above the nature of this scientific discipline necessarily means that our definition of AI will change over time. Instead of considering AI as a discipline that can be clearly delineated, with uncomplicated definitions and fixed methodologies, it is more useful to see it as a complex and diverse field focused on a certain horizon. The dot on that horizon is the understanding and simulation of all human intellectual skills. This goal is also called 'artificial general intelligence' or AGI (other names are 'strong AI' and 'full AI'). However, it remains to be seen whether this dot, with such a generic definition of AI, will ever be reached. Most experts believe that this it at least several decades away – if it is ever attained at all.[13]

A fixed definition of AI as the imitation of full human intelligence is of little use for the purposes of this report. We need a definition that captures the whole range of applications finding their way into practice today and in the near future. The definition from the AI HLEG provides the necessary freedom of scope. Describing AI as "systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals", this encompasses all the applications we currently qualify as AI and at the same time provides scope for future changes to that qualification. Alongside advanced machine learning and deep learning technologies, this definition also allows for other technologies, including the more traditional approaches mentioned above, as used by many government bodies. In short, this definition is sufficiently strict to distinguish AI from algorithms and digital technology in general, while at the same time open enough to include future developments. Figure 2.1 provides an overview of the definitions discussed and the AI HLEG definition used in this report.

[13] Martin Ford (2018) interviewed 23 experts for his book *Architects of Intelligence: The Truth about AI from the People Building It* and asked them, 'What year do you think human-level AI might be achieved, with a 50% probability?' Most were only willing to respond anonymously and the year they suggested, on average, was 2099 – so almost 80 years from now. We will return to the potential of AGI in later chapters.



AI is ...

"The use of algorithms. The term 'algorithm' refers to a specific instruction for solving a problem or performing a calculation."

"The imitation of all human intellectual abilities by computers."

"The imitation of various complex human skills by machines."

"Technology that can function appropriately and with foresight in its environment."

"Systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals."

Fig. 2.1 Various definitions of AI

It is worth emphasizing that the current applications considered as AI according to this definition all fall under the heading 'narrow' or 'weak' AI.[14] The AI that we are familiar with today focuses on specific skills, such as image or speech recognition, and has little to do with the full spectrum of human cognitive capabilities covered by AGI. This does not alter the fact that current AI applications can and do give rise to major issues, too. The American professor of Machine Learning Pedro Domingos has put this nicely; in his view we focus too much on a future AGI and too little on the narrow AI that is already all around us. "People worry that computers will get too smart and take over the world," he says, "but the real problem is that they're too stupid and they've already taken over the world."[15]

The fact that AI is difficult to define is linked to the evolution of this discipline. We now take a closer look at how that evolution took place. A short historical overview is not only relevant as a background for understanding AI, it is also the prelude to the next chapter in which we see that AI has reached a turning point.

2.2 AI Prior to the Lab

It is possible to date the birth of some disciplines very precisely. AI is one. Its conception in the laboratory is often dated to 1956, during a summer school at Dartmouth College in New Hampshire, USA. AI did not come out of the blue, however. The

[14] With regard to the terms 'narrow AI' and 'weak AI', we prefer the former. The latter obviously suggests that this type of AI lacks strength, whereas that may well not be the case. In fact, it is simply limited to a well-defined (read: 'narrow') domain. For example, a computer program may be very good at translating texts but still 'narrow' because it cannot be used for image recognition.

[15] Domingos, 2017: 286.

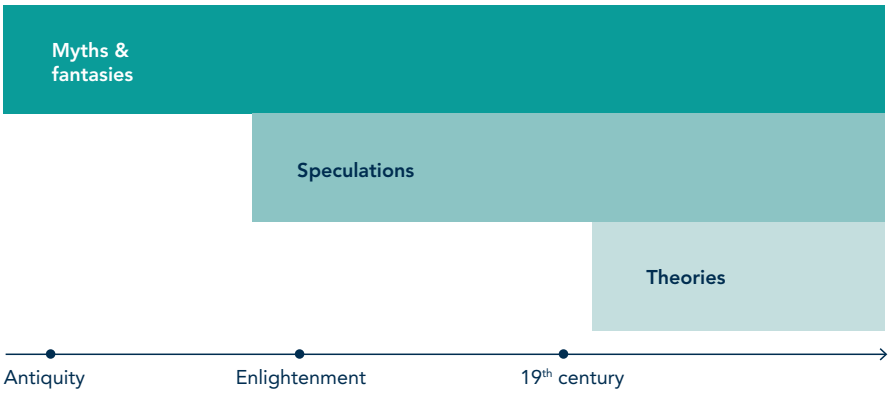


Fig. 2.2 Three phases of AI prior to the lab

technology already had a long history before it was first seriously investigated as a scientific discipline.

This history can be divided roughly into three phases: early mythical representations of artificial forms of life and intelligence; speculations about thinking machines during the Enlightenment; and the establishment of the theoretical foundation for the computer (see Fig. 2.2). The latter was the springboard for the development of AI as a separate discipline. We now discuss these three phases in turn, but bearing in mind that in practice they have never been mutually exclusive. Myths have always existed and there has always been creative speculation about the future in parallel with the theoretical research into AI. Nevertheless, the phases reveal how the nature and focus of AI thinking have changed over time.

2.2.1 The Mythical Representation of AI

Myths and stories about what we would now call AI have been around for centuries (see Fig. 2.3). The ancient Greeks in particular celebrated a multitude of characters in their mythology who can be characterized as artificial forms of intelligence.[16] Take Talos, a robot created by the great inventor Daedalus to protect the island of Crete. Every day, Talos would run circles around the island and throw stones at any approaching ships he spotted. This is clearly a myth about a mechanical super-soldier. A robotic exoskeleton used by the US Army now bears the same name.

Daedalus, the ancient world's great inventor, is famous for the wings that cost the life of his son Icarus, but he was also the inventor of all manner of artificial intelligence, such as moving statues as well as Talos. According to the myth, this robot was eventually defeated by the witch Medea, who tricked it into disabling itself. So, while Daedalus was an AI inventor, in the same legend Medea was able to magically

[16] In *Gods and Robots – Myths, Machines, and Ancient Dreams of Technology*, Mayor (2018) examines the phenomenon of 'made, not born' in antiquity.

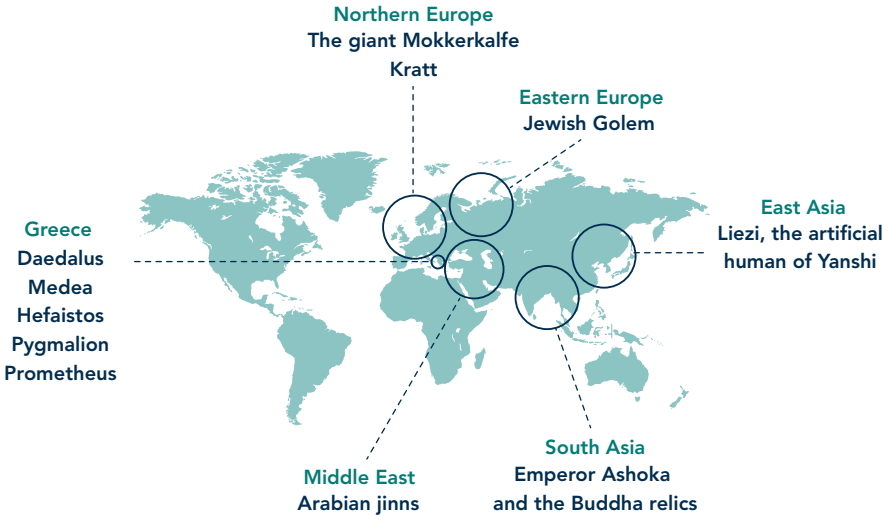


Fig. 2.3 Ancient myths about AI

control his AI. Moreover, her father was responsible for creating artificial soldiers who could fight without needing rest.

In addition to the two human characters of Daedalus and Medea, various Greek gods were also associated with artificial intelligence. Hephaistos, the blacksmith of the gods, was assisted in his workshop by mechanical helpers. He also built tools that moved independently and a heavenly gate that opened automatically. The titan Prometheus 'built' humans and stole fire from the gods for them. To punish humankind, Zeus created a kind of robot, the mechanical woman Pandora, who poured out all kinds of suffering on humans when she opened her jar ('Pandora's box'). A less grim example is the myth of Pygmalion. A sculptor, he fell in love with a statue he had made, upon which Aphrodite brought it to life and he made his creation, named Galatea, his wife. So the ancient Greeks were already imagining what we now would call killer robots, mechanical assistants and sex robots in their mythology.

There are also stories about forms of AI in other traditions, such as the Jewish golem and the mythical jinn (genies) of Arabia who can grant wishes. The Buddhist story *Lokapannatti* tells how the emperor Ashoka wanted to lay his hands on the relics of the Buddha, which were protected by dangerous mechanical guards made in Rome.[17] Norse mythology tells of the giant Hrungnir, built to battle Thor. The *Liezi*, an ancient Chinese text, relates the story of the craftsman Yan Shi, who built an automaton with leather for muscles and wood for bones.[18] Estonia has a legend about the Kratt, a magical creature made of hay and household items that did everything its owner asked. If the Kratt was not kept busy, it became a danger to its owner. The modern law in Estonia that governs liability for the use of algorithms is known there as the 'Kratt Law'.

[17] Zarkadakis, 2015: 34.

[18] Brynjolfsson & McAfee, 2014: 250.

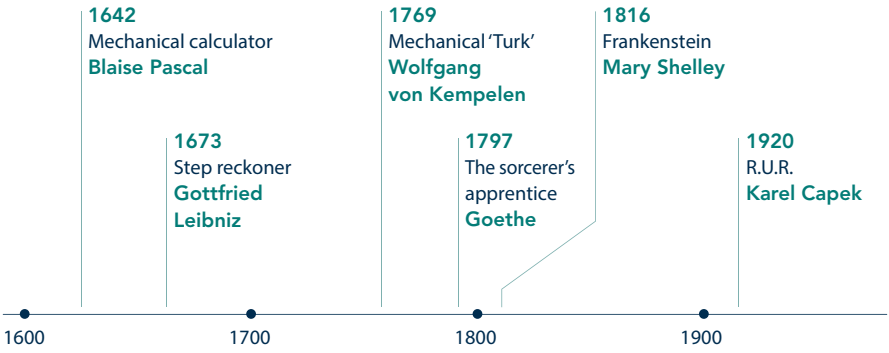


Fig. 2.4 Timeline of speculations about AI

2.2.2 Speculation About Thinking Machines

The next phase was heralded by the 'mechanization of the world'[19] envisaged in the work of thinkers like Galileo Galilei, Isaac Newton and René Descartes. Their mechanical worldview was accompanied by the construction of all kinds of novel machines. Artificial intelligence was still far beyond the realm of possibility, but the new devices did lead to speculation about its creation (see Fig. 2.4) – speculation that was no longer mythical, but mechanical in nature.

In 1642 Blaise Pascal built a mechanical calculator which he said was "closer to thought than anything done by animals".[20] Gottfried Leibniz constructed an instrument he called the 'step reckoner' in 1673, which could be used to perform arithmetical calculations. This laid the foundation for many future computers.[21] The philosophers of the time speculated about such devices using the term 'automata'.

In 1769 Wolfgang von Kempelen built a highly sophisticated machine – or so people long thought. He gained worldwide fame after offering his mechanical 'Turk' to the Austrian Empress Maria Theresa. The huge device was an automatic chess machine, which toured the western world for 48 years and defeated opponents like Napoleon Bonaparte and Benjamin Franklin. It was not until the 1820s that it was discovered to be a total fake: there was a man inside the machine moving the pieces.[22] As an aside, the company Amazon has a platform called Mechanical Turk where people can arrange to have tasks done cheaply online. While more open than Von Kempen's original, here too the work is done by people behind the scenes we do not see.

Speculation about AI could also take magical forms during this period. Goethe's story of the sorcerer's apprentice, made famous in Disney's animated film *Fantasia* starring Mickey Mouse, is about an apprentice who uses a spell to make a broom

[19] Described by Dijksterhuis in *De mechanisering van het wereldbeeld* ('The mechanization of the world view', 1950).

[20] Russell, 2019: 40.

[21] Broussard, 2019: 76.

[22] Zarkadakis, 2015: 37.

fetch water. When it turns out he does not know the spell to make the process stop, and instead the broom begins to multiply itself, a disaster unfolds that only ends when the wizard returns.[23] Other magical stories about phenomena similar to AI include *Pinocchio* and the horror story by W. W. Jacobs about a monkey's paw that grants three wishes with terrible consequences.

Tales of magic have also spilled over into stories a little closer to scientific reality, in the form of science fiction. In 1816 a group of writers meeting near Geneva was forced to spend long periods indoors because of a volcanic eruption in what is now Indonesia. That caused the so-called 'Year Without a Summer', when abnormal rainfall kept people inside. Inspired by the magical stories of E. T. A. Hoffman, Lord Byron suggested that each member of the group write a supernatural story, upon which Mary Shelley penned the first version of her famous novel *Frankenstein*.[24]

The story of a scientist who creates an artificial form of life that ultimately turns against its creator has become the archetype of the risks of modern technology. This motif lives on in countless films, including classics like *Blade Runner* (1982), *The Terminator* (1984) and *The Matrix* (1999).

Another important work of literary science fiction in the context of speculation about AI is *R.U.R.* by the Czech author Karel Capek. It is in this book that the writer introduces the term 'robot', a word derived from the Old Church Slavonic word 'rabota', meaning corvée or forced labour. This story also reveals a classic fear of AI; in it the artificial labourers ('roboti') created in a factory rebel against their creators and ultimately destroy humankind.[25] Capek's book was published in 1920, by which time the next phase – much more concrete thinking about AI – had long since begun.

2.2.3 The Theory of AI

From the second half of the nineteenth century onwards, the idea of AI as 'thinking computers' became less fantastical and entered the realm of serious theoretical consideration (see Fig. 2.5). This development occurred in parallel with the theorization and construction of the first computers.

Ada Lovelace – daughter of the poet Byron, instigator of the writing session that had produced *Frankenstein* – would play an important role in this field in the 1840s. She envisaged a machine that could play complex music based on logic, and also advance scientific research in general. Her acquaintance Charles Babbage designed such a device in 1834 and called it the 'Analytical Engine'.[26] He had earlier failed in his efforts to build an enormously complex Difference Engine and so instead created the Analytical Engine as an alternative with which he hoped to construct mathematical and astronomical tables.[27] Lovelace, however, saw a much wider use for a

[23] Wiener, 1964: 57.

[24] Zarkadakis, 2015: 60–63.

[25] Rid, 2016: 83.

[26] Boden, 2018: 6.

[27] Freeman & Louçã, 2001: 309.

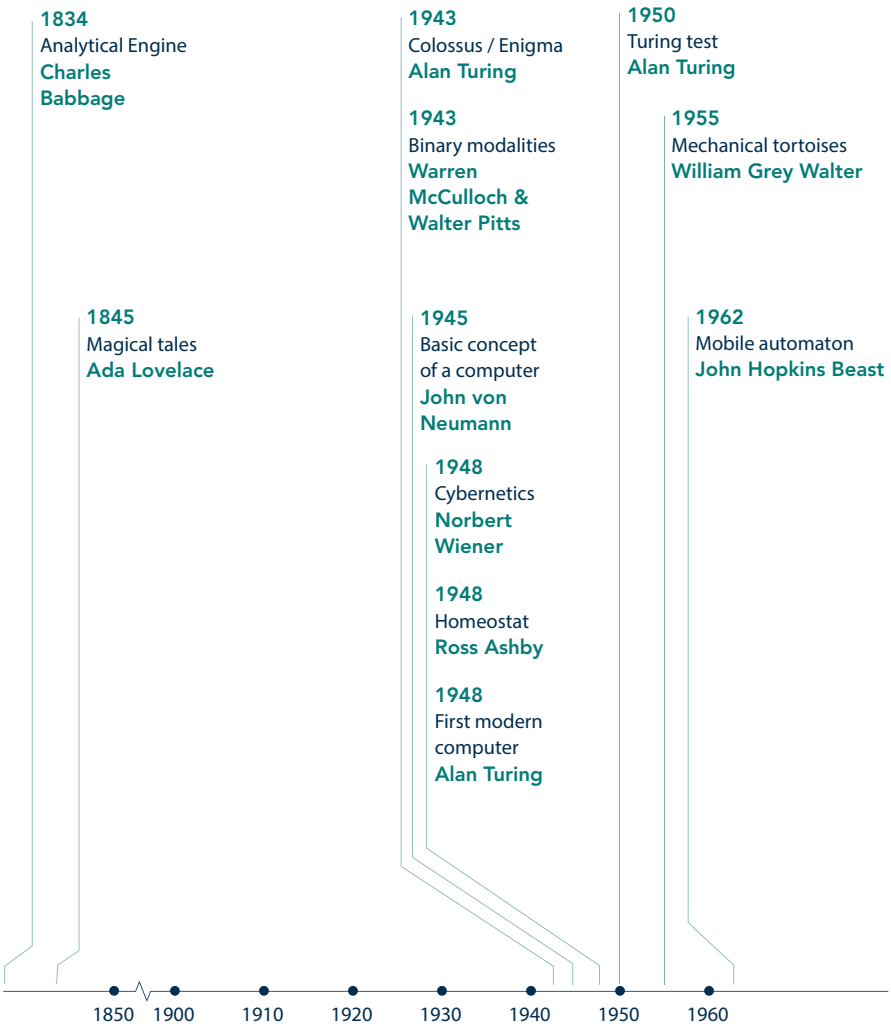


Fig. 2.5 Timeline of theories of AI

'thinking machine' that could reason about "all the subjects in the universe".[28] She even wrote programs for the hypothetical device. However, science at that time was not advanced enough to actually build such computers.

That point would not be reached until the Second World War, when computing power was needed to defend against air raids. The use of fast-moving planes to drop bombs made it impossible for the human operators of anti-aircraft systems to respond quickly enough when relying on their eyesight alone. Instead, their targets' trajectories needed to be calculated mathematically. Research in that field laid the foundations for the modern computer and for another discipline that would emerge

[28] Russell, 2019: 40.

in the 1950s, cybernetics. This work immediately raised questions about automation and human control that are still relevant today.

"The time factor has become so narrow for all operators," a military spokesperson said at the time, "that the human link, which seems to be the only immutable factor in the whole problem and which is particularly fickle, has increasingly become the weakest link in the chain of operations, such that it has become clear that this link must be removed from the sequence."[29]

The development of the computer was given another boost during the war by the British research programme Colossus, which aimed to crack the Nazis' secret communication system known as Enigma. One of the leading lights in this top-secret project at Bletchley Park was Alan Turing, often regarded as the father of both computers and AI. He went on to help develop the first truly modern computer in Manchester in 1948. Two years after that, in 1950, he wrote a paper proposing a thought experiment in the form of an 'imitation game' for a computer pretending to be a human being.[30] This has come to be known as the Turing test. A computer passes if a human is unable to establish that its written answers to their questions were provided by a person or a computer. Variants of this test are still used, for example, to compare AI systems with human abilities such as recognizing images or using language.[31]

Another important theoretical contribution to this field was a paper by psychiatrist and neurologist Warren McCulloch and mathematician Walter Pitts.[32] In this they combined Turing's work on computers with Bertrand Russell's propositional logic and Charles Sherrington's theory of neural synapses. Their most important contribution was that they demonstrated binary modalities (a situation with two options) in various domains and thus developed a common language for neurophysiology, logic and computation. The distinction between 'true and false' in logic was now linked to the 'on or off' state of neurons and the computer values '0 and 1' in Turing machines.[33]

John von Neumann continued to develop the basic concept of a computer with components such as the central processor, memory and input-output devices.[34] Another important founder of AI theory was Norbert Wiener. He coined the term 'cybernetics' in 1948 to describe "the study of control and communication in

[29] Rid, 2016: 37–38.

[30] Turing, 2009 [1950].

[31] There has also been criticism of the use of language in the Turing test. Yann LeCun, a prominent AI scientist, suggested in an interview that there are forms of intelligence that have nothing to do with language (Ford, 2018: 129). Some animals, for example, use less complex language than humans but still form good models of the world and can employ tools.

[32] McCulloch & Pitts, 1943.

[33] In a lecture at Yale in the 1950s, the scientist John von Neumann described the similarity between the computer and the brain as follows: "The nervous pulses can clearly be viewed as (two-valued) markers, in the sense discussed previously: the absence of a pulse represents one value (say, the binary digit 0), and the presence of one represents the other (say, the binary digit 1)." von Neumann, 2012 [1958]: 43.

[34] Freeman & Louçã, 2001: 310.

animals and machines".[35] The key idea was that people, animals and machines could all be understood according to a number of basic principles. The first of these is control: all those entities strive to counter entropy and to control their environment using the principle of 'feedback', which is the "ability to adapt future behaviour to past experience". Through the mechanism of continuous adjustment and feedback, organisms and machines ensure that equilibrium, or homeostasis, is achieved. Wiener used thermostats and servomechanisms as metaphors to explain these processes. Although cybernetics did not last long as a separate scientific field, its core concepts now permeate all manner of disciplines (Box 2.1).[36]

Thanks to such advances, during this period scientists were ready to stop just dreaming and thinking about AI and start actually developing the technology and experimenting with it in the laboratory. The starting gun for this race was fired in 1956.

Key Points: AI Prior to the Lab

- Mythical representations of AI have been around for centuries.
- The most celebrated examples are the ancient Greek stories about Daedalus, Medea, Hephaistos, Prometheus and Pygmalion.
- The mechanization of the world view from the seventeenth century onwards made the construction of all kinds of machines possible. This went hand in hand with speculation about mechanical brains.
- Fictional stories about artificial intelligence appeared from the Industrial Revolution onwards, including *Frankenstein* and *R.U.R.*
- The theoretical foundations for AI were laid when the first computers were built by people like Alan Turing.

Box 2.1: The Homeostat and Electronic Tortoises

In 1948 the Briton Ross Ashby unveiled his 'homeostat', a machine able to hold four electromagnets in a stable position. In that same year *The Herald* wrote of this 'protobrain' that "the clicking brain is cleverer than man's".[37] Another highlight of the cybernetics movement in the 1950s was William Grey Walter's electronic tortoises. These small devices could walk around without bumping into obstacles and locate where in the room their charger was if their battery was weak. Moreover, they also exhibited complex social behaviour as a group. A later example of a cybernetic machine was the John Hopkins Beast, which in the early 1960s was able to trundle through corridors using sonar and a photocell eye to find a charging point.[38]

[35] Wiener, 2019 [1965].

[36] Rid, 2016: 47–52. Famous cyberneticians in various disciplines include the neurophysiologist Warren McCulloch, the physicist Heinz von Foerster, the management theorist Stafford Beer, the philosopher Humberto Maturana, the political scientist Karl Deutsch, the anthropologist Gregory Bateson and the sociologist Talcott Parsons.

[37] Rid, 2016: 53–55.

[38] Moravec, 1988: 7.

2.3 AI in the Lab

2.3.1 The First Wave

As mentioned previously, the beginnings of AI as a discipline can be dated very precisely.[39] After all the myths, speculation and theorizing, artificial intelligence appeared in a lab for the first time in 1956 when a group of scientists made it the subject of a specific event: the Dartmouth Summer Research Project on Artificial Intelligence. This was a six-week brainstorming gathering attended by several of the discipline's founders. The organizers were very optimistic about what they could achieve with this group in a few weeks, as is evident from the proposal they wrote to the Rockefeller Foundation.

> We propose … a 2-month, 10-man study of artificial intelligence … The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.[40]

The proposal was overambitious, and research is still being carried out today in all the areas it mentioned. With this project, however, these scientists formulated a research agenda that launched AI as a discipline.

The summer project was organized by John McCarthy and Marvin Minsky. It was McCarthy who coined the term 'artificial intelligence' in 1956. Minsky was a leading figure in the history of AI and over the years came to be involved in many prominent high-tech projects around the world. The two men also established the Artificial Intelligence Lab at MIT. This was later renamed the MIT Media Lab and is still a centre for the creative use of new technology.[41] Among those present at the summer project were Herbert Simon (Nobel laureate in Economics and winner of the Turing Award, responsible for the idea of 'bounded rationality', amongst other things, and founder of the Carnegie Institute of Technology), John Nash (mathematician, game theorist and another Nobel laureate in Economics) and Arthur Samuel (pioneer of computer games and the man credited with popularizing the term 'machine learning'). These leading scientists were responsible for bringing AI to the lab.

This landmark event heralded a period of great optimism and broad interest in the field of AI, which has come to be known as the first 'AI spring' (or 'wave'). Various programs were developed that could play the board game draughts (checkers), although none was very good yet. The version developed by Samuel did eventually succeed in defeating its human creator, which caused a stir, although he was not

[39] The history of a scientific discipline can be written in several ways. It can focus on the fundamental science, for instance, or on practical inventions and applications. One example is the difference between the development of the natural sciences and the inventions of the Industrial Revolution. In this chapter we combine both perspectives, but the idea of waves in AI is rooted mainly in that of inventions and applications.

[40] Bostrom, 2016: 6.

[41] Broussard, 2019: 69–70.

known as a great player of the game. Wiener wrote in 1964 that, while Samuel was eventually able to beat the program again after some instruction, "the method of its learning was no different in principle from that of the human being who learns to play checkers". He also expected that the same would happen with chess in ten to twenty-five years, and that people would lose interest in both games as a consequence.[42]

Exciting breakthroughs followed when AI systems began focusing on a different category of challenges: logical and conceptual problems. For example, a 'Logic Theory Machine' was built to prove Bertrand Russell's logical theorems. It not only succeeded in proving eighteen of them, it also developed a more elegant proof of one. This was important because, while Samuel was a mediocre draughts player, Bertrand Russell was a leading logician.

The next milestone was the 'General Problem Solver'. This was a program that could, in principle, be applied to solve any problem – hence the name. By translating problems into goals, subgoals, actions and operators, the software could then reason what the right answer was. One example of a problem it solved is the classic logical puzzle of the river crossing.[43]

By the mid-1960s the first students of the AI pioneers were working on programs that could prove geometric theorems and successfully complete intelligence tests, maths problems and calculus exams. So, the discipline was making progress, but its impact outside the lab was very limited. There were some interesting experiments with robots, as in the late 1960s at the Stanford Research Institute; its Shakey the Robot was able to find its way about through reasoning.[44] The American technology company General Electric built impressive robots such as the Beetle and an exoskeleton that enabled humans to lift heavy weights.[45] These robots were not very practical, though.

At the same time, there were grand expectations of AI. In 1965 Herbert Simon predicted that "machines will be capable, within twenty years, of doing any work a man can do".[46] Meanwhile, the British mathematician Irving Jack Good foresaw a machine-induced 'intelligence explosion'. This would also be the last invention of humankind, because machines would now be the most intelligent beings on earth and therefore do all the inventing.[47]

AI caught the imagination of people outside science as well. In 1967 the computer program MacHack VI was made an honorary member of the American Chess Federation, despite having won very few matches.[48] A few years later the film *Colossus: The Forbin Project* was released. In this a computer program is handed control of the US military

[42] Wiener, 1964: 22–24. It would eventually take thirty years for a computer to defeat a chess grandmaster, as we shall see shortly. In any case, people have not lost their interest in these games since sophisticated programs have learned to play them.

[43] Boden, 2018: 10. In this logical problem, three entities all have to cross a river. Only two can cross at the same time. Each entity threatens to harm one of the others, so not every duo can cross together. The problem is: which combinations can be formed to convey everyone to the other side unharmed?

[44] Russell, 2019: 52.

[45] Rid, 2016: 136.

[46] Brynjolfsson & McAfee, 2014: 141.

[47] Rid, 2016: 148. The writer Vernor Vinge would later coin the term 'singularity' for this scenario.

[48] Bakker & Korsten, 2021: 24.

arsenal because it can make better decisions than humans and is unhindered by emotions. After the Soviets reveal a similar project, the two programs start communicating with one another – but in a way that is incomprehensible to their human creators – and subsequently take control of the entire world. Their pre-programmed goal of world peace is achieved, but the price is the freedom of the human race.

This gap between hopeful expectations and harsh reality did not go unnoticed, and from the second half of the 1960s onwards there was increasing criticism of AI research. The philosopher Hubert Dreyfus would remain critical of the potential of AI throughout his life. In 1965 he wrote a study called *AI and Alchemy*, commissioned by the Rand Corporation (the think tank of the American armed forces), in which he concluded that intelligent machines would not be developed any time in the near future. In a 1966 report to the US government, the Automatic Language Processing Advisory Committee concluded that little progress had been made. The National Research Council subsequently phased out its funding of AI. In the United Kingdom, Sir James Lighthill was commissioned in 1973 to conduct a survey of the topic; this brought to light considerable criticism of its failure to achieve the grandiose goals that had been promised. As a result, a lot of research funding was withdrawn in the UK as well.[49]

One problem encountered by many AI systems at this time was the so-called 'combinatorial explosion'. These systems solved problems by exploring all possible options, but they quickly reached the limits of their computing power when dealing with huge numbers of possible combinations. More heuristic approaches, based on rules of thumb, were needed to reduce the number of combinations. However, these did not yet exist. This and other problems – such as the lack of data to feed the systems and the limited capacity of the hardware – meant that progress with AI stalled.

Meanwhile, its practical applications were also proving unreliable. When an AI system was developed during the Cold War, in the 1960s, to translate Russian communications, the results proved less than impressive. One famous example was its translation of "the spirit is willing, but the flesh is weak" as "the vodka is good, but the meat is rotten".[50] During the course of the 1970s, the earlier optimism turned to pessimism. There were too few breakthroughs, so criticism of AI grew, and funding dried up. The first 'AI winter' had set in and put an end to its first wave. Figure 2.6 provides an overview of the emergence of AI as a scientific discipline.

2.3.2 Two Approaches

It is important to note that two distinct approaches to AI gained particular prominence during this first wave. While it is true that there were others as well (we will explain these later), these two still dominate the field to this day. The first is 'rule-based', also known as 'symbolic' or 'logical', AI (along with other names) and emerged in the 1970s in the form of so-called 'expert systems'. Its core principle is

[49] Leung, 2019: 253.

[50] Russell & Norvig, 2021: 21.

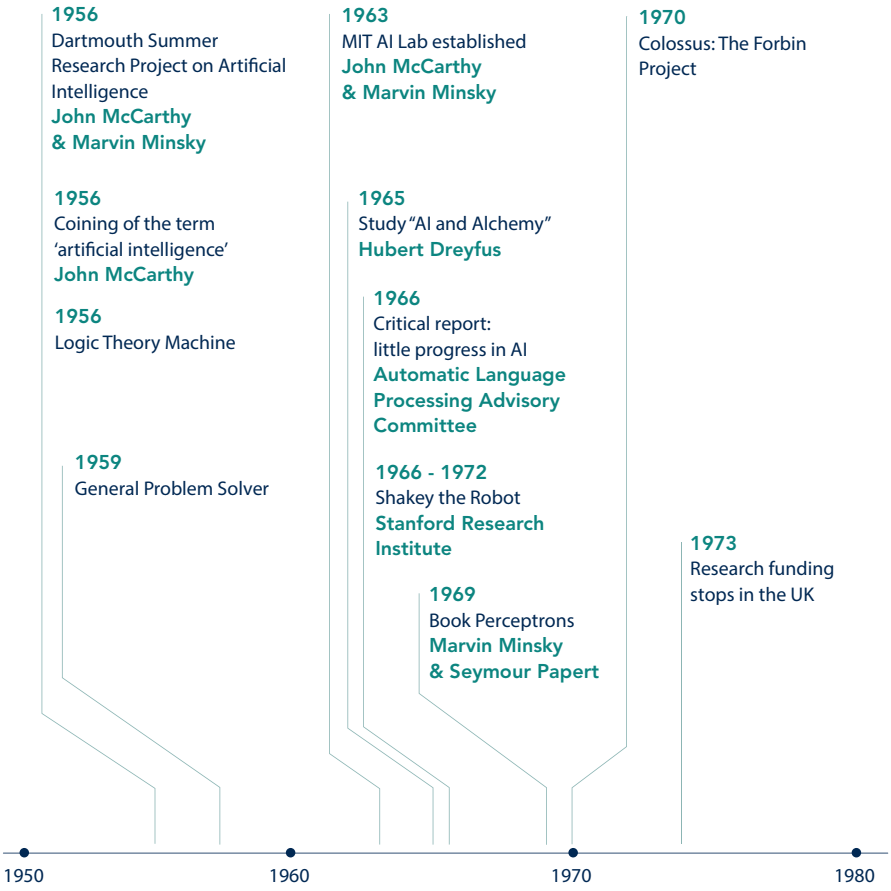


Fig. 2.6 Timeline of the emergence of AI as a discipline (first wave)

that computers learn by encoding logical rules with formulas of the type 'IF X, THEN Y'. The use of logic and rules is also why the term 'symbolic AI' is used, as this approach follows rules that can be expressed in human symbols.

The second approach uses artificial neural networks (ANNs) and is also called 'connectionism'. This includes the deep learning and parallel distributed processing methods that have received a lot of attention in recent years. The central idea here is to simulate the functioning of neurons in the human brain. For this purpose, sets of 'artificial neurons' are built into networks that can receive and send information. These networks are then fed with large amounts of data and try to distil patterns from it. In this case the rules are not drawn up by humans in advance. Most ANNs are based on a principle formulated as early as 1949 by Donald Hebb, a Canadian psychologist, in his book *The Organization of Behaviour*: "Neurons that fire together, wire together".[51] In other words, if two neurons are frequently activated at the same time, they become connected.

[51] Domingos, 2017: 93.

Both approaches to AI were there from the start. While many of the founding fathers at the 1956 summer school followed the rule-based approach, the first artificial neuron was also created around the same time at Cornell University.[52] The difference can be explained as follows. To be able to recognize a cat in a photo, in the first approach a series of 'IF-THEN' rules are established: the presence of certain colours, a given number of limbs, certain facial forms, whiskers, etc., means that it is a cat. With these rules, a program can 'reason' what the data means.

In the second approach, the program might be presented a large number of photos labelled as 'cat' and 'non-cat'. The program distils patterns based on this data, which it then uses to recognize the presence of a cat in subsequent photos. Rather than using labels, another variant of this approach instead presents large numbers of images and then allows the program to come up with its own clustering of cats. In both variants, however, it is not the rules programmed by people, but the patterns identified by the program that determine the outcome.

As already noted, both approaches were explored during the first AI wave. One example of an application of neural networks was Frank Rosenblatt's 'perceptron', an algorithm he invented which learned to recognize letters without these being preprogrammed. This was attracted much media interest in the 1960s. Symbolic AI, however, remained dominant. The Logical Theory Machine and General Problem Solver mentioned earlier were both examples of systems within this strand. For decades it would remain the dominant approach within AI.

The proponents of symbolic AI also expressed much criticism of neural networks. They considered that approach unreliable and of limited use due to its lack of rules. In 1969 Marvin Minsky, an ardent supporter of the symbolic approach, wrote a book called *Perceptrons* with Seymour Papert. This amounted to a painstaking critique of the neural network approach, backed by examples of mathematical proofs of problems it could not solve. To many this appeared to sound the death knell for that approach.[53] Such criticism not only marginalized the position of neural networks, it also contributed towards the onset of the first AI winter.

2.3.3 The Second Wave

In 1982 *Time* magazine named the personal computer its Man of the Year. This coincided with a revival of interest in AI, and the discipline entered a second spring. At the time, the programming language Prolog was used for many logical reasoning systems. In 1982 the Japanese government invested a huge sum in a Prolog-based AI system in the form of the Fifth-Generation Computer Systems Project.[54] This was a far-reaching, ten-year partnership between the government and industry and

[52] Greenfield, 2017: 214.

[53] From an interview with Geoffrey Hinton (Ford, 2018: 83).

[54] Russell, 2019: 271.

was intended to boost the discipline in Japan by establishing a 'parallel computing architecture'. At a time when there was widespread fear of Japanese economic growth, several Western countries quickly followed suit with their own projects.

To keep up with the competition, the US established the Microelectronics and Computer Technology Corporation (MCC), a research consortium. In 1984 MCC's principal scientist, Douglas Lenat, launched a huge project called Cyc. Initiated with the full support of Marvin Minsky, this is still running today and involves collecting vast amounts of human knowledge about how the world works.[55] In 1983 DARPA, the scientific arm of the US Department of Defense, announced a Strategic Computing Initiative (SCI) that would invest one billion dollars in the field over ten years.[56] Both the Japanese and the American research projects took a broad approach to AI, with hardware and human interfaces also playing an important role, for example.[57] In 1983 the United Kingdom announced its response to the Japanese plans in the form of the Alvey Programme.

One important development during this second wave was the emergence in the 1970s of expert systems within symbolic AI. These are a form of rule-based AI where human experts in a particular domain are asked to formulate the rules for a program. One example was MYCIN, a program trained by medical experts to help doctors identify infectious diseases and prescribe appropriate medication. The Dendral project involved the analysis of molecules in organic chemistry. Expert systems were also developed to plan manufacturing processes and solve complex mathematical problems; for example, the Macsyma project. Such systems thus found practical applications outside the lab.

Some were developed in the Netherlands, too, in the 1980s and tested in pilot projects. These addressed themes including the implementation of social security and criminal sentencing policies.[58] In part thanks to specific research programmes and funding provided by the Dutch Research Council (Nederlandse Organisatie voor Wetenschappelijk Onderzoek, NWO) and various universities, but also by a number of government departments, the Netherlands was even able to establish an international profile with a relatively large research community in the field of legal knowledge-based systems. An important early facilitator in this respect was JURIX, the Foundation for Legal Knowledge-Based Systems, an organization of 'legal tech' researchers from the Netherlands and Flanders. It has held annual international conferences since 1988; their proceedings – all available online – testify to the rich Dutch and Flemish academic history of research on and development of AI applications in the legal domain.[59] Another prominent platform is the Benelux Association for Artificial Intelligence (Benelux Vereniging voor Kunstmatige Intelligentie, BNVKI), originally formed in the Netherlands in 1981 (as the NVKI) but later

[55] Domingos, 2017: 35.

[56] Leung, 2019: 254.

[57] Russell & Norvig, 2020: 24.

[58] Hage & Verheij, 1999.

[59] www.jurix.nl/proceedings/

connecting scientists from Belgium and Luxembourg as well. The US Office for Technology Assessment has called expert systems "the first real commercial products of about 25 years of AI research"[60] and in 1984 the front page of *The New York Times* reported that they held out "the prospect of computer-aided decisions based on more wisdom than any one person can contain".[61]

Nevertheless, the results of this second wave were ultimately disappointing. The big ambitions of the major national projects were never achieved, either in Japan, the US or Europe. Their poor results were why the US SCI drastically scaled down its funding. Among the problems to limit the potential of these projects were hardware issues. This period culminated with the bankruptcy of several specialized companies in the field in the late 1980s.[62] But the expert systems also had their own problems. They tended to be highly complex, so minor errors in the rules had disastrous consequences for the results and systems could fail when two rules contradicted each other.[63] The Cyc project is still ongoing but has failed to live up to expectations throughout almost four decades of existence.[64] By the late 1980s, therefore, another AI winter had set in: the second wave had run out of momentum.

2.3.4 The Third Wave

In the 1990s, however, AI again began to attract attention and eventually flourish anew. Initially, the logical systems approach had several successes. One of the most iconic of these was the victory of IBM's Deep Blue program over chess grandmaster Garry Kasparov, in 1997. At the time this was considered a fundamental breakthrough. The successor to that program, named Watson, later participated in the US television quiz show *Jeopardy!*, in which contestants have to formulate questions to match given answers. In 2011 Watson defeated the game's reigning human champions. This was seen as proof that AI was approaching mastery of human language, another major breakthrough. Both cases are examples of the use of symbolic AI, in which the lessons of chess masters and answers from previous players of *Jeopardy!* were fed to the programs as rules. At the same time, however, experts were becoming increasingly dissatisfied with this approach.

---

[60] Leung, 2019: 259.

[61] Dreyfus & Dreyfus, 1986: ix.

[62] Leung, 2019: 255.

[63] The idea of expressing the limits of human behaviour and language in rules had been explored earlier by philosophers such as Ludwig Wittgenstein (Wittgenstein, 1984).

[64] According to Ray Kurzweil, a proponent of neural networks, Cyc has actually achieved almost nothing (Ford, 2018: 233). That, however, is an oversimplification. Such projects form the foundations of techniques such as knowledge graphs, which are now important for the functioning of search engines like Google. This also demonstrates why the two approaches are not mutually exclusive and in practice often go hand in hand.

Although both events were huge landmarks in the eyes of the public, in reality the truth was more prosaic. Stuart Russell describes how the foundations of chess algorithms were laid by Claude Shannon in 1950, with further innovations following in the 1960s. Thereafter, these programs improved according to a predictable pattern, in parallel with the growth of computing power. This was easily measurable against the scores recorded by human chess players. The linear pattern predicted that the score of a grandmaster would be achieved in the 1990s – exactly when Deep Blue defeated Kasparov. So that was not so much a breakthrough as a milestone that had been anticipated as part of a predictable pattern.[65] Deep Blue won by brute force, thanks to its superior computing power. Moreover, various chess champions had fed heuristic principles into its software. Instead of the smart computer beating the human, this victory could also be seen as the triumph of a collective comprising a computer program and numerous human players over a single grandmaster.[66] It was man and machine together that were superior to a human opponent.

The computer's victory in *Jeopardy!* is also questionable. It would be incorrect to claim that the program could understand the complex natural language of humans. The game has a very formalized question-and-answer design, and many of the questions can be found on a typical Wikipedia page. This makes them relatively easy to answer for a program that can rapidly search mountains of information for keywords; that does not require an in-depth understanding of language.

While these logical systems only began to attract attention in the 1990s, other forms of AI had been making progress for far longer and the momentum eventually shifted towards the neural network approach. This trend had already begun in the mid-1980s when fundamental research into the so-called 'backpropagation algorithm' (in which multiple layers of neural networks are trained) improved the process of pattern recognition. At about the same time the US Department of Defense recognized that its funding programme had been unfairly neglecting the neural networks approach. Under the banner of 'parallel distributed processing', neural networks returned to centre stage in 1986. In a book published the previous year, John Haugeland had introduced the term GOFAI ('good old-fashioned AI') – a phrase which has since become a pejorative term for symbolic AI. In the same period Judea Pearl began applying probability theory rather than logical reasoning to AI.

Breakthroughs below the radar were thus undermining the dominant rule-based approach. A paper on backpropagation was rejected for a leading AI conference in the early 1980s and, according to Yann LeCun, researchers at the time even used code words to mask the fact that they were working with neural networks.[67] It took time for the importance of this new approach to become recognized. For example, Jeff Hawkins said in 2004 that AI had fewer skills than a mouse when it came to image recognition.[68]

[65] Russell, 2019: 62–63.

[66] Ihde, 2010.

[67] From an interview with Yann LeCun (Ford, 2018: 122).

[68] Tegmark, 2017: 79.

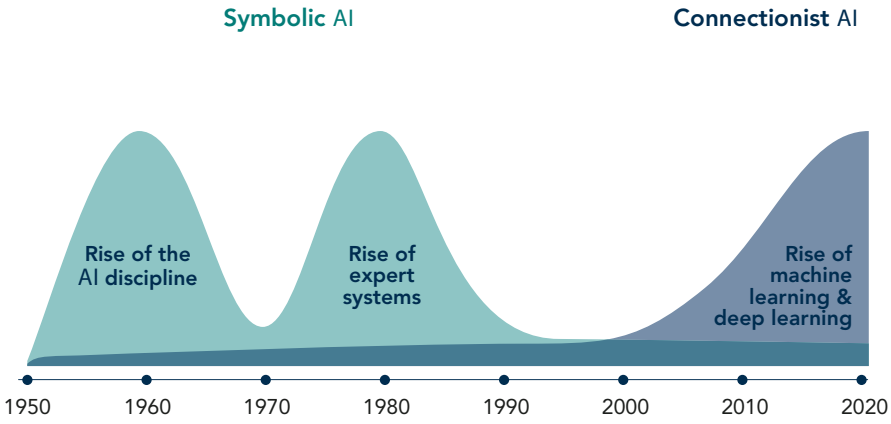


Fig. 2.7 The transition from a symbolic to a connectionist AI

At that time, it was thought it would take another century before a computer could beat a human in the Asian game go, which has many more combinations of moves than chess.[69] In fact, Google's AlphaGo program defeated world champion Lee Sedol in 2016. This was made possible thanks to recent breakthroughs in the approach to neural networks, in which researchers such as Yann LeCun and Andrew Ng played an important role. But it is Geoffrey Hinton who is often seen as the father of those advances. Together with David Rumelhart and Ronald Williams, he had already popularized the use of the backpropagation algorithm in a paper published in *Nature* in 1986. That algorithm traces the contribution made by the output layer back to hidden layers behind it, where individual units are identified that need to be modified to make the algorithm work more effectively. For a long time, the 'backprop' had only a single hidden layer, but more have recently been distinguished. Backpropagation thus addresses a central problem of ANNs: the representation of hierarchy. Relationships can now be distinguished at different levels and the success factors of the algorithm are also determined at all levels (called 'credit assignment').[70] Such neural networks have since been used, for instance, to simulate the price of shares on the stock exchange. Figure 2.7 shows the historical development of the two approaches to AI.

In 1989 Yann LeCun applied backprop to train neural networks to recognize handwritten postcodes. He used convolutional neural networks (CNNs), where complex images are broken down into smaller parts to make image recognition

[69] Tonin, 2019: 1.

[70] In the book *Perceptrons*, which was highly critical of the neural networks approach, Minsky and Papert demonstrated that it was unable to solve the problem of the 'exclusive OR' (XOR). But Rumelhart, Hinton and Williams showed that backpropagation could learn XOR.

Box 2.2: Three Forms of Machine Learning
ML can be subdivided into three different forms: supervised, unsupervised and reinforcement learning. In supervised learning, a program is fed data with labels as in our earlier example of 'cat' versus 'non-cat'. The algorithm is trained on that input and then tested to see if it can correctly apply the labels to new data.

Unsupervised learning has no training step and so the algorithm needs to search for patterns within the data by itself. It is fed large amounts of unlabelled data, in which it starts to recognize patterns of its own accord. The starting point here is that clusters of characteristics in the data will also form clusters in the future. Supervised learning is ideal when it is clear what is being searched for. If the researchers themselves are not yet sure what patterns are hidden within data and are curious to know what they are, then unsupervised learning is the more appropriate method.

more efficient. This was another important contribution to contemporary AI programs.[71]

In another paper, written in 2012, Hinton introduced the idea of 'dropout', which addresses the specific problem of 'overfitting' in neural network training. That occurs when a model focuses so strongly on training with existing data that it cannot effectively process new information. Hinton's work gave an enormous boost to the applicability of neural networks in the field of machine learning. The use of multiple layers in the training process is why it is called 'deep' learning; each layer provides a more complex representation of the input based on the previous one. For example, while the first layer may be able to identify corners and dots, the second one can distinguish parts of a face such as the tip of a nose or the iris of an eye. The third layer can recognize whole noses and eyes, and so it goes on until you reach a layer that recognizes the face of an individual person (Box 2.2).[72]

The third form is applicable in other contexts, such as playing a game. Here it is not about giving a right or wrong answer or clustering data, but about strategies that can ultimately lead to winning or losing. In these cases, the reinforcement learning approach is more suitable. The algorithm is trained by rewarding it for following certain strategies. In recent years reinforcement learning has been applied to various classic computer games such as Pacman and the Atari portfolio, as well as to 'normal' card games and poker. The algorithm is given the goal of optimizing the value of the score and then correlates all kinds of actions with that score to develop an optimum strategy.

In 2012 Hinton's team won an international competition in the field of 'computer vision' – image processing using AI. They achieved a margin of error of 16%,

[71] Marcus & Davi, 2019: 52.
[72] Domingos, 2017: 117.

whereas no team before them had ever managed less than 25%. A few years earlier the same team had been successful in using neural networks for speech recognition after a demonstration by two students in Toronto. But the gains in computer vision in 2012 were the real revelation for many researchers.[73] Deep learning proliferated, and in 2017 almost all the teams in the competition could boast margins of error lower than 5% – comparable with human scores. That improvement continues to this day. The application of DL has since gained momentum, with the scientific breakthroughs using neural networks prompting an explosion of activity in this approach to AI. We are currently at the height of this latest AI summer. In the next chapter we look in more detail at the developments that has set in motion outside the lab: in the market and in wider society.

It is clear that the rapid expansion of AI in recent years has its origins in fundamental scientific research. Big companies like Google have subsequently rushed to hire talented researchers in this field, but it is scientists at universities who have been responsible for the most important breakthroughs.

In addition to these academic milestones, two other factors underlie the recent rise and application of AI. The first is the growth in processing power, as encapsulated in Moore's Law. This pattern, that the number of transistors on a chip roughly doubles every two years, has been observed consistently in the computer industry for decades. It means that more and more computing power is becoming available while prices continue to fall. Hence the fact that the smartphones of today surpass the computing power of the very best computers of only a few decades ago. We noted earlier how the first 'AI winter' was caused in part by the combinatorial explosion. The increase in computing power provided the solution to this problem. A further leap in that power came from the chip industry, using graphic processing units (GPUs) rather than the classic central processing units (CPUs). GPUs were originally developed for complex graphics in the gaming industry but were subsequently found to enable many more parallel calculations in AI systems as well.[74] Since 2015, tensor processing units (TPUs) specifically designed for ML applications have also come into use.

The other factor that has contributed to the current AI wave is the increase in the amount of data. This is closely linked to the rise of the internet. In the past algorithms could only be applied to a limited range of data sources. In recent decades, however, as people have started to use the internet more and more, and directly and indirectly to generate a lot more digital information, the amount of data available for AI systems to analyse has increased significantly.

The 'digital breadcrumbs' we leave behind on the internet are now food for training AI algorithms. But we are helping with this training in other ways, too. By tagging personal names in photos on Facebook, for example, people provide algorithms with labels that can be used to train facial recognition software. One specific dataset that is very important for this kind of training is ImageNet, an open database of

[73] From an interview with Geoffrey Hinton (Ford, 2018: 77).

[74] Kelly, 2017: 38.

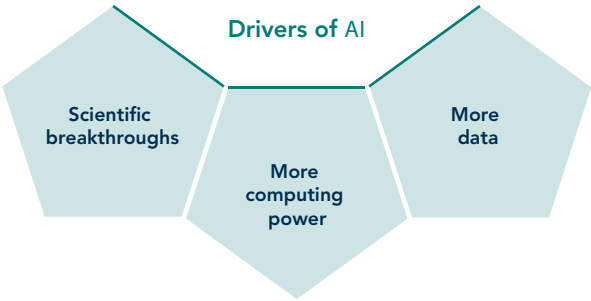


Fig. 2.8 Three drivers of progress in AI

more than 14 million hand-labelled images. The 'internet of things' (the growing number of sensors and connections in the physical environment) is also contributing to the growth in data.

The triad of scientific breakthroughs, greater computing power and more data has allowed AI to take off in a big way recently (see Fig. 2.8). As mentioned, this expansion has been driven mostly by the application of machine learning as part of the neural network approach, and within ML by the development of deep learning.

Key Points: AI in the Lab

- In the lab AI has ridden three waves of development. Between these were two 'winters' when scientific progress ground to a halt, hardware capacity was inadequate, and expectations were not met.
- The first wave began with the Dartmouth Summer Research Project in 1956. At that time AI was used mainly for games such as draughts, in early robots and to solve mathematical problems. Two further waves, dominated by progress in symbolic AI and then neural networks, would follow.
- The second wave began in the 1980s, driven in part by the international competition between Japan, the US and Europe. This produced expert systems, the first major commercial applications of AI.
- The third wave began in the 1990s with major achievements in symbolic AI, but only properly gained momentum some years later due to advances in the field of machine learning and its subfield of deep learning. The scientific breakthroughs in this area, together with increases in computing power and data volumes, are the driving force behind this wave, which continues to this day.

References

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: The simple economics of Artificial Intelligence*. Harvard Business Press.

Bakker, S., & Korsten, P. (2021). *Artificiële Intelligentie Als Een general purpose technology: Strategische Belangen Van Verantwoorde Inzet In Historisch Perspectief*

(WRR Working Paper nr. 41). Wetenschappelijke Raad voor het Regeringsbeleid. Available at: https://www.wrr.nl/publicaties/working-papers/2021/02/16/artificiele-intelligentie-als-een-general-purpose-technology

Boden, M. (2018). *Artificial Intelligence: A very short introduction*. Oxford University Press.

Bostrom, N. (2016). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Broussard, M. (2019). *Artificial Unintelligence: How computers misunderstand the world*. MIT Press.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton and Company.

DenkWerk. (2018). *Artificial Intelligence in Nederland: Zelf Aan Het Stuur*. Available at: https://denkwerk.online/media/1029/artitificial_intelligence_in_nederland_juli_2018.pdf

Dennett, D. (2019). What can we do? In J. Brockman (red.), *Possible minds: Twenty-five ways of looking at AI* (pp. 41–53). Penguin.

Dignum, V. (2019). *Responsible Artificial Intelligence: How to develop and use AI in a responsible way*. Springer.

Domingos, P. (2017). *The master algorithm: How the Quest for the ultimate learning machine will remake our world*. Penguin Random House.

Dreyfus, H., & Dreyfus, S. (1986). *Mind over Machine*. The Free Press.

Floridi, L. (2014). *The Fourth Revolution: How the Infosphere is reshaping human reality*. Oxford University Press.

Ford, M. (2018). *Architects of Intelligence*. Packt Publishing.

Freeman, C., & Louçã, F. (2001). *As time Goes By: From the industrial revolutions to the information revolution*. Oxford University Press.

Greenfield, A. (2017). *Radical technologies: The design of everyday life*. Verso Books.

Hage, J., & Verheij, B. (1999). Rechtsinformatica: De Stand Van Zaken In De Wetenschap. In A. Oskamp and A. Lodder (reds.), *Informatietechnologie voor juristen. Handboek voor de jurist in de 21e eeuw* (pp. 65–92). Kluwer.

High-Level Expert Group on Artificial Intelligence. (2019). *A definition of AI: Main capabilities and scientific disciplines*. European Commission. Available at: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56341

Ihde, D. (2010). *Embodied technics*. Automatic Press/vip.

Kelly, K. (2017). *The Inevitable: Understanding the 12 technological forces that will shape our future*. Penguin.

Leung, J. (2019). *Who will govern Artificial Intelligence? Learning from the history of strategic politics in emerging technologies*. Dissertation, Oxford University. Available at: https://ora.ox.ac.uk/objects/uuid:ea3c7cb8-2464-45f1-a47c-c7b568f27665

Marcus, G., & Davi, E. (2019). *Rebooting AI: Building Artificial Intelligence we can trust*. Vintage.

Mayor, A. (2018). *Gods and Robots: Myths, machines, and ancient dreams of technology*. Princeton University Press.

McCulloch, W., & Pitts, W. (1943). A Logical Calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics, 5*(4), 115–133.

Moravec, H. (1988). *Mind Children: The future of robot and human intelligence*. Harvard University Press.

Nilsson, N. (2009). *The Quest for Artificial Intelligence*. Cambridge University Press.

Rid, T. (2016). *Rise of the machines: A cybernetic history*. WW Norton & Company.

Russell, S. (2019). *Human compatible: Artificial Intelligence and the problem of control*. Penguin.

Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A modern approach* (4th ed.). Pearson.

Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A modern approach*. Pearson.

Tegmark, M. (2017). *Life 3.0: Being Human in the age of Artificial Intelligence*. Penguin.

Tonin, M. (2019). Artificial Intelligence: Implications for NATO's Armed Forces. *149 stctts 19 E rev. 1 fin.*

Turing, A. (2009 [1950]). Computing machinery and Intelligence. In R. Epstein, G. Roberts, and G. Beber (reds.), *Parsing the turing test*. Springer.

von Neumann, J. (2012 [1958]). *The Computer and the Brain.* Yale University Press.
Wiener, N. (1964). *God and Golem, Inc.: A comment on certain points where cybernetics impinges on religion*. MIT Press.
Wiener (2019 [1965]). Cybernetics: Or control and communication in the animal and the machine. MIT Press.
Wittgenstein, L. (1984). *Tractatus logico-philosophicus. Tagebücher 1914–1916. Philosophische Untersuchungen*. Suhrkamp.
Zarkadakis, G. (2015). *In our own image: Will artificial intelligence save or destroy us?* Ebury Publishing.