MODULE 1 UNIT 1.1

The History of AI

Ver. 1.0

Table of contents (M1U1)

1. Introduction

This first unit of the notes will explore the history of AI from the 1950s to the present. As you read through this history, take note of the moments and milestones that are most likely to have had an impact on the development of the eight definitions you explored in this section.

The term “artificial intelligence” (AI) was officially coined by John McCarthy in 1956, and has since grown into a fascinating field of science and engineering (Russell & Norvig, 2016:1). Google CEO Sundar Pichai is of the opinion that “AI is one of the most important things humanity is working on [and] it is more profound than… electricity or fire” (Clifford, 2018). Although optimism prevails, prominent voices like Elon Musk and the late Stephen Hawking have also warned that AI may pose a threat to the human race (Sulleyman, 2017; Cellan-Jones, 2014). While Hawking described AI as “either the best, or the worst thing, ever to happen to humanity”, Musk emphasized the importance of AI regulation stating that “there should be some regulatory oversight, maybe at the national and international level” to ensure that AI does not become a risk to humanity in the future (Hern, 2016; Marr, 2017).

Whatever the long-term impact of AI may be, these statements clearly demonstrate the immense transformative power that this new technology holds.

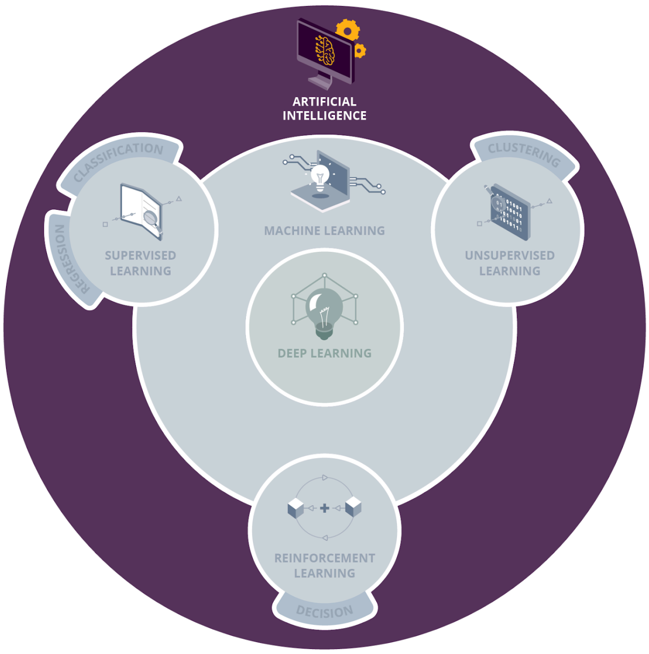

Figure 1.1.1: The programme framework - AI (AI Programme, Oxford, 2025).

The aim of this module is to give you a sound understanding of AI as a general concept. In later modules, you will delve deeper into the various other concepts illustrated in the framework to further understand how each of them relates to AI.

This set of notes will elaborate on the definition of AI, and then follow with an in-depth discussion of the history of AI from 1950 to the present day, and a quick look at the current development of the technology.

2. The Timeline of AI

Since the early 1950s, AI has been an active scientific discipline associated with widespread research and breakthroughs, with modern AI reaching a point that is far beyond the anticipation of early AI pioneers (Wooldridge, 2018:4). Even though it is an exciting field in science, some people may be sceptical about AI and what it can achieve, and this is not entirely unfounded (Wooldridge, 2018:4). Throughout the history of AI, there have been a number of instances where researchers have oversold its potential.

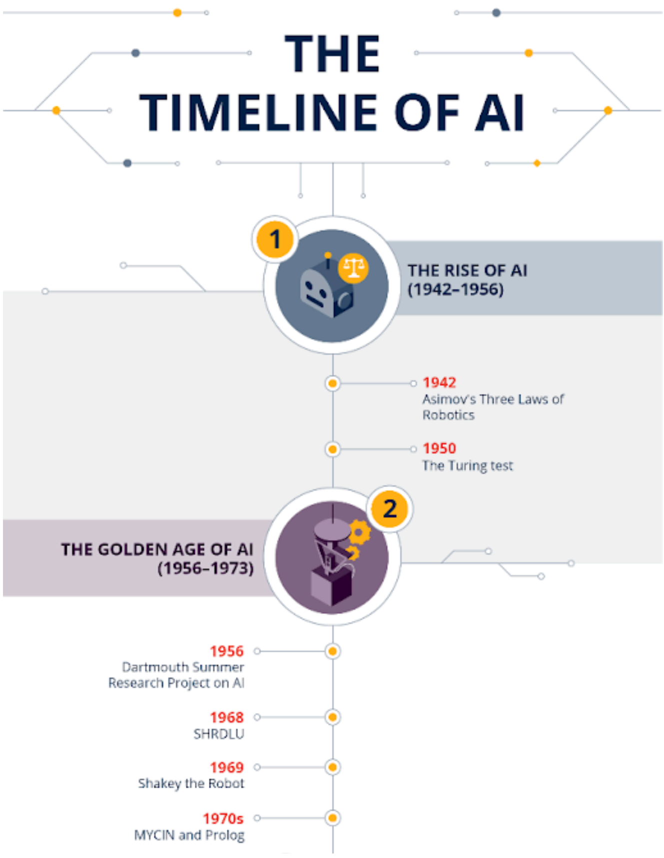

Figure 1.1.2: The timeline of AI (part 1).

The resulting excitement and optimism have often led to unrealistic predictions of what AI will be able to achieve in the future (Wooldridge, 2018:4).

The rise of AI (1942-1956)

Asimov's Three Laws of Robotics (1942) and The Turing test (1950)

Isaac Asimov was an American writer, and one of many people who had envisioned that machines may one day be intelligent. His work inspired scientists to think differently about robotics and the recreation of intelligence in machines (Biography, 2019).
He published a short story titled "Runaround" in Astounding Fiction magazine in 1942, where he first laid out his famous Three Laws of Robotics, which, in his stories, are encoded into robots to prevent them from turning on their creators (BBC, n.d.). The three laws developed by Asimov are:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given to it by human beings, except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Explore Deeper (Recommended)

For those who wish to venture beyond the main learning path, this section offers a deeper dive into fascinating sub-topics. These materials are optional, but they provide a richer context and a more holistic perspective on the world of Gen AI.

- The following **video** captures Isaac Asimov reciting the Three Laws of Robotics he invented to ensure that the machine never turns on the creator.
- In opposition to the fantasy laws created by Asimov, in his 2019 **book** – Human Compatible – Stuart Russell writes about how AI is mainly concerned with reaching its objectives, programmed by humans, but without the nuance and care of humans. An example could be an AI generated delivery bot which has the objective to deliver food in the fastest time. The bot may go to extreme measures to make sure the food is delivered as soon as possible, but may not take into account the safety of cars or pedestrians it might come in contact with. Read more about Russell's theories on the current state of AI and its potential risks.

In 1950, another mark in history was made with the publication of the Turing test by Alan Turing. Although his theory was not without fault, it did spark excitement for AI and gave rise to intensified AI research, specifically in natural language processing (NLP) (Wooldridge, 2018:6).

The “golden age” of AI (1956-1973)

What is commonly known as the golden age of AI started in 1956 and was characterised by a heightened enthusiasm about the possibilities of the field (Wooldridge, 2018:6).

Dartmouth Summer (1956)

As you would have read in Wooldridge’s (2018:12) book, the term “artificial intelligence” was coined in the year 1956 by John McCarthy during a summer conference at Dartmouth University. At this conference, intellectuals debated the topic of AI and how to approach it. The two main opinions argued at this conference are represented in Figure 4 (BBC, n.d.

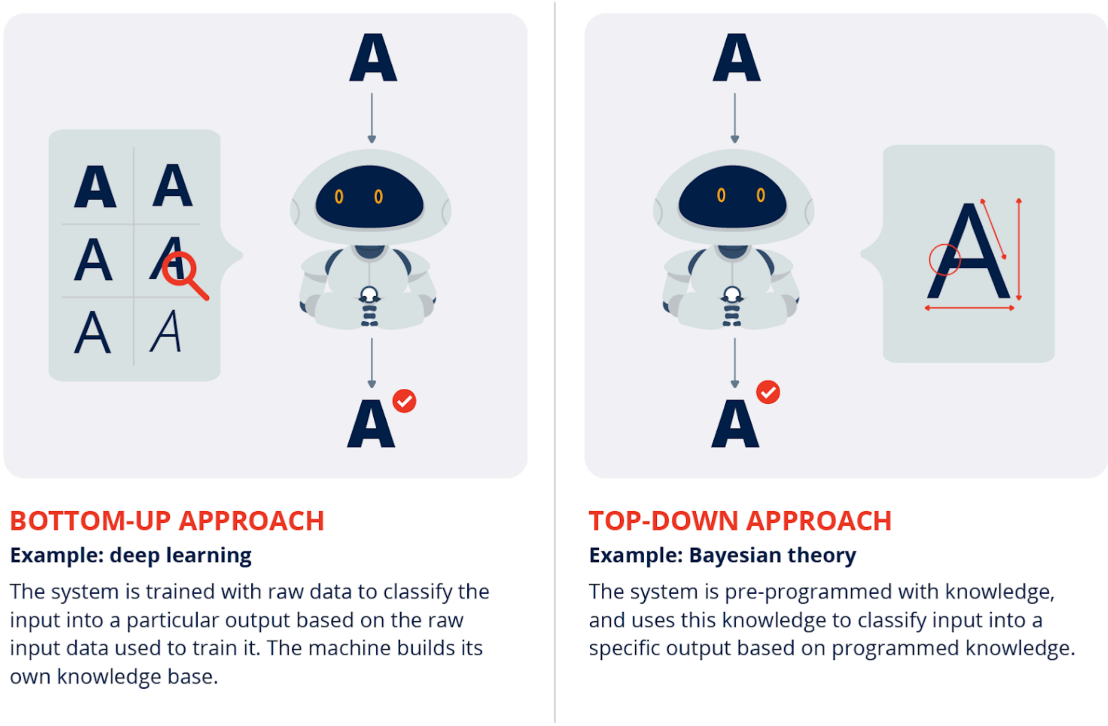

Figure 1.1.3: The differences between top-down & bottom-up approaches in AI research.

After much debate and deliberation between the proponents of these two approaches, the top-down approach emerged as the dominant view of those at the conference. Professor Marvin Minsky’s contribution to this debate led to the United States government providing funding for AI research in the hope of discovering technology that could prove valuable in the Cold War (BBC, n.d.). Minsky was also responsible for a so-called change in AI, specifically relating to the excitement and potential of neural networks. The change that took place resulted from Minsky’s book titled Perceptrons, where he highlighted the issues associated with neural networks in AI that had to be overcome to make the innovation in this field more useful. This publication by Minsky ultimately halted research in the field of neural networks for many years (Knight, 2016).

SHRDLU and the Blocks World(1968)

Various achievements were celebrated in the golden age, including a system named SHRDLU, based on the "Blocks World". The SHRDLU, initially released in 1968, represented a breakthrough moment in the history of AI because it was initially believed to have the potential to understand the English language (Wooldridge, 2018:18).

Shakey the Robot (1969)

In 1966, Shakey the Robot made its appearance. Shakey, also called the "first electronic person", was the first robot able to move around autonomously by perceiving its surroundings. It had the ability to plan and find routes, and rearrange small objects in its way. It was fitted with a camera, a range finder, and bump sensors, and used programs to process the data to navigate through a given area (SRI International, n.d.).

The invention of Shakey sparked excitement and public imagination, and it was elected to the Carnegie Mellon's Robot Hall of Fame in 2004 (SRI International, n.d.).

Even though Shakey represented a breakthrough in that it was the first general-purpose mobile robot, able to observe its surroundings and make decisions accordingly, it was extremely slow. When faced with moving objects in its surroundings, Shakey would often stop for an hour at a time, updating its map and planning the next move (BBC, n.d.). Shakey served as a painful confirmation of the extent to which AI was lagging behind the predictions made in the initial stages of the golden age (BBC, n.d.).

Explore Deeper (Recommended)

For those who wish to venture beyond the main learning path, this section offers a deeper dive into fascinating sub-topics. These materials are optional, but they provide a richer context and a more holistic perspective on the world of Gen AI.

- To learn more about Shakey the Robot watch this video featuring some of the team members responsible for its development

.

MYCIN - Rule-based expert system (early 1970s)

During the 1970s, various other systems were applied in industry, including MYCIN and Prolog, which added to the excitement around AI and its potential (Smith et al., 2006:12).

(add more from the book on "MYCIN", "CYC: The ultimate expert system?"

The MYCIN system was intended to be a doctor's assistant, providing expert advice about blood diseases. It was a classic example of a rule-based expert system, and one of the most famous systems from the days of knowledge-based AI.

MYCIN's blood disease knowledge was represented in about 200 rules. Each rule captured a discrete piece of knowledge about blood diseases, carefully written after discussions with human experts. This task -to extract knowledge from human experts- is difficult and time-consuming because people often can't articulate the knowledge they have.

MYCIN pioneered many features that became essential for expert systems, including: to explain a conclusion in a way humans could understand it; and to be able to cope with uncertainty.

MYCIN performed better in its area of expertise than human specialists - this was an impressive and much lauded observation. However, its expertise was very narrow.

Expert systems represented a new stage in the evolution of AI, shifting from its initial emphasis on general problem-solvers to expressing in-code human reasoning, i.e., drawing inferences and arriving at logical conclusions. The new focus was on knowledge, specifically the knowledge of specialized (narrow) domain experts and specifically, their heuristic knowledge.

PROLOG - Logic programming language(early 1970s)

Prolog, short for "Programming in Logic", is a logic programming language that was widely used in the early days of Artificial Intelligence (AI) research and development. It was designed to enable symbolic reasoning and knowledge representation, making it suitable for tasks like theorem proving, expert systems, and natural language processing.

Prolog has its roots in first-order logic, a formal logic. Unlike many other programming languages, Prolog is intended primarily as a declarative programming language: the program is a set of facts and rules, which define relations. A computation is initiated by running a query over the program.

Prolog was instrumental in developing early AI systems, particularly in areas like expert systems (which provide expert-level advice) and natural language processing (understanding and generating human language).

In essence, Prolog provided a powerful framework for building AI systems that could reason logically and represent knowledge in a structured way, making it a significant part of AI's early history and still relevant in some specialized applications today.

The complexity barrier

During the golden age, scientists tried to explain why systems like the Blocks World did not seem to work in real-world scenarios. It was during this time that the mathematical theory “computational complexity” was published, which offered an explanation as to why real-world implementations were unsuccessful (Wooldridge, 2018:8).

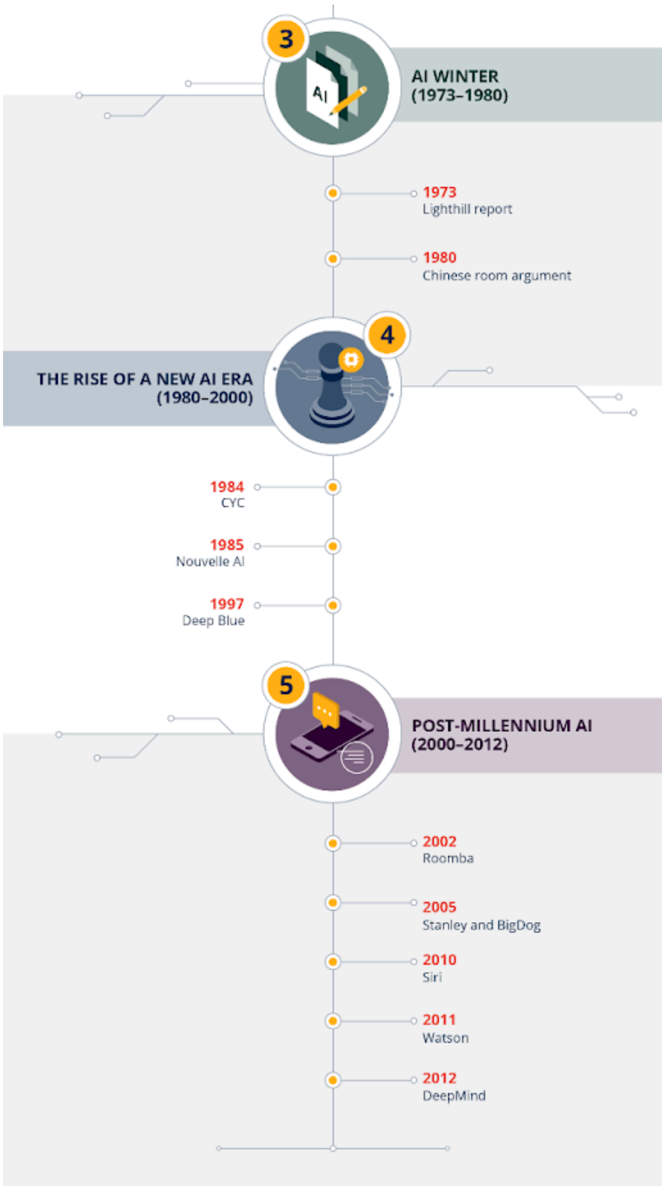

Figure 1.1.4: The timeline of AI (part 2).

The AI "winter" (1973-1980)

As a result of Minsky's Perceptrons, the complexity barrier, and the failure of AI in real-world scenarios, funding was retracted for AI research, which led to the period known as the AI winter (Wooldridge, 2018:22). AI was in trouble, and by 1973, dire reports were given on the state of AI in the United Kingdom (BBC, n.d.).

The Lighthill report (1973)

In the Lighthill report, Professor Sir James Lighthill argued that machines will always remain at a level of an "experienced amateur" and that facial recognition will forever remain beyond their reach (BBC, n.d.).

Explore Deeper (Recommended)

For those who wish to venture beyond the main learning path, this section offers a deeper dive into fascinating sub-topics. These materials are optional, but they provide a richer context and a more holistic perspective on the world of Gen AI.

- To learn more about the main points of Professor Lighthill's criticism, read the full Lighthill report.

{By contrasting these criticisms with the current achievements of AI, it is important to note that the report makes a clear distinction between the automation of actual tasks (the Category A) where AI now excels in many cases, while also highlighting the problems related to the more fundamental research — trying to develop a true artificial intelligence by replicating human central nerval systems.}

The Chinese room argument (1980)

The Chinese room argument, proposed by philosopher John Searle, is a thought experiment designed to challenge the idea that a computer can truly understand language simply by executing a program that mimics human language understanding. It argues that while a computer might appear to understand a language (like Chinese) by following a set of rules (a program), it doesn't actually possess genuine understanding or consciousness in the way humans do.

The Thought Experiment:

1. The Setup:

Imagine a person who doesn’t speak Chinese locked inside a room. They are given a rule book (in English) and a pile of Chinese symbols.

2. The Input:

People outside the room slide Chinese sentences under the door.

3. The Process:

The person inside the room follows the English instructions in the rule book to manipulate the Chinese symbols. The instructions might say things like, "If you see symbol 'X', then output symbol 'Y'," and so on.

4. The Output:

The person inside the room passes out the manipulated Chinese symbols as a response, according to the rules.

5. The Illusion:

To the people outside, it appears that someone inside the room understands Chinese and is having a conversation. However, the person inside the room has no understanding of the meaning of the Chinese symbols; they are simply following instructions.

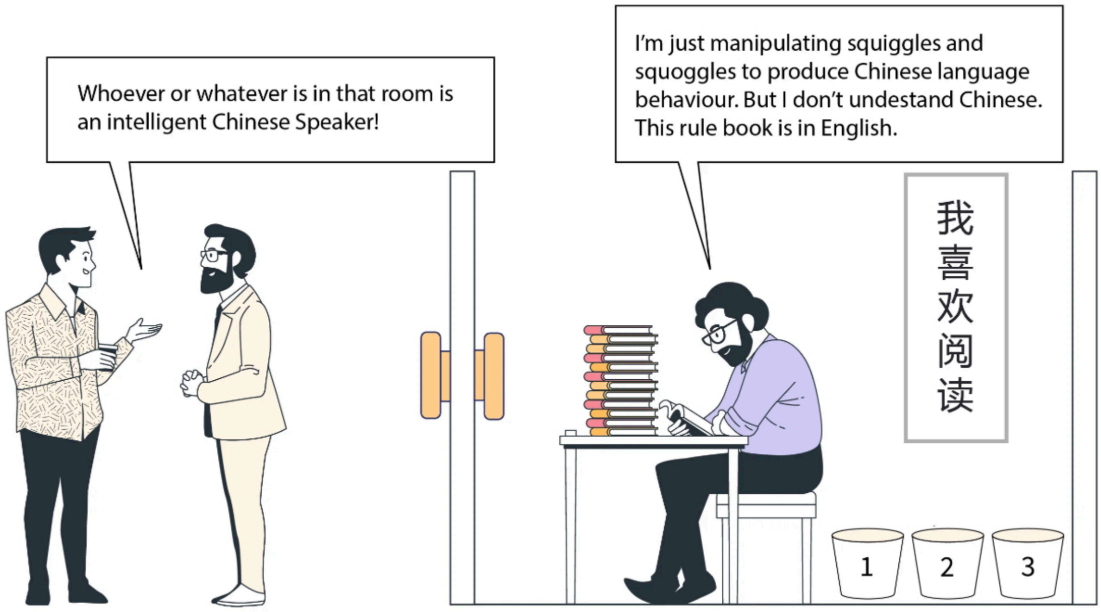

Figure 1.1.5: The Chinese room argument.

The argument highlights the difference between syntax (the rules for manipulating symbols) and semantics (the meaning of symbols). The computer (or the person in the room) manipulates symbols syntactically, but it doesn't grasp the semantics or meaning of those symbols.

Searle's argument is aimed at refuting "strong AI," the idea that a computer program can possess genuine understanding and consciousness simply by simulating human cognitive processes. He argues that even if a computer program can pass the Turing test (appearing to understand language), it doesn't necessarily mean it understands anything.

The Chinese room argument raises questions about the possibility of achieving Artificial General Intelligence (AGI), which aims to create AI systems with human-level intelligence and understanding. It suggests that simply mimicking human-like behavior might not be enough to achieve true understanding.

The Rise of a new AI era (1980-2000)

In the early 1980s, the AI winter came to an end when a new-found optimism for AI and the value it held for businesses emerged. The potential for commercial value attracted investors to reinvest in AI research (BBC, n.d.). Research was shifted from general AI to focus more on narrow AI, which focuses specifically on solving one particular problem (BBC, n.d.).

CYC (1984)

In 1984, Doug Lenat initiated arguably the boldest experiment from that time – a project known as CYC, which is still ongoing (Wooldridge, 2018:26). The vision was to create a computer program that had common sense, by importing facts, rules, and knowledge into the program to enable it to process natural language conversations and eventually be able to read text (Stork & Clarke, 1998).

Nouvelle AI and the robotics revolution(1985)

By the mid-1980s, Nouvelle AI and the robotics revolution started, driven by Rodney Brooks, who had a vision to create robots that are intelligent (Wooldridge, 2016:32). This was followed by the idea of behavioural AI, which started being developed in the late 1980s.

Deep Blue (1997)

Following intensive research, AI came of age in 1997 when IBM invented the machine, Deep Blue, that ended up victorious in a game of chess against the then world champion, Garry Kasparov (BBC, n.d.). This sparked new excitement as well as controversy in the field of AI as it moved into the new millennium.

Explore Deeper (Recommended)

For those who wish to venture beyond the main learning path, this section offers a deeper dive into fascinating sub-topics. These materials are optional, but they provide a richer context and a more holistic perspective on the world of Gen AI.

- Watch the following video if you would like to learn more about Deep Blue and the chess game it won against Garry Kasparov.

Post-millenium AI (2000-2011)

Roomba (2002)

In 2002, robotics-based advocate and scientist, Rodney Brooks, created the Roomba vacuum cleaner – the first robot used in the home (BBC, n.d.). The Roomba is fitted with intelligent sensors to navigate through a home, or any other space, for the purpose of cleaning the floors (iRobot Corporation, 2019). Although the Roomba seemed insignificant in relation to the ambitions of AI researchers, it paved the way for a new era of autonomous robots with the ability to successfully complete specific tasks (BBC, n.d.).

Explore Deeper (Recommended)

For those who wish to venture beyond the main learning path, this section offers a deeper dive into fascinating sub-topics. These materials are optional, but they provide a richer context and a more holistic perspective on the world of Gen AI.

- Watch the following video to find out more about how the Roomba works.

Stanley and BigDog (2004-05)

The year 2004 heralded the dawn of the autonomous vehicle when DARPA (a US military defence research agency) launched a competition, challenging researchers to enter their autonomous vehicles into their Grand Challenge. The first Grand Challenge had no victors, which led to DARPA holding a second Grand Challenge the next year. It was during this challenge that the autonomous vehicle and robot, STANLEY, won the race, outperforming 195 other entrants (Wooldridge, 2018:36). Since then, a lot of research was done – and is still being done – to improve autonomous vehicles.

In 2005, the US military realised the potential of AI in warfare, and invested in a research project of Boston Dynamics. One of these projects, named BigDog, is a robot that was developed to walk alongside troops to carry supplies in rough terrain, which is inaccessible for conventional vehicles (BBC, n.d.; Greenmeier, 2008). These robots, that are the size of a large dog or a small mule, rely on four legs that enable it to absorb shock (Greenmeier, 2008; Boston Dynamics, n.d.). By design, BigDog has the ability to maintain its balance in rough terrain and, as is evident in Video 9, can even do so when attacked or hit by something or someone while in motion. BigDog can run at a speed of about 6mph, climb angles of up to 35 degrees, walk in snow and water, and carry a load of up to 150kg (Boston Dynamics, n.d.).

Explore Deeper (Recommended)

For those who wish to venture beyond the main learning path, this section offers a deeper dive into fascinating sub-topics. These materials are optional, but they provide a richer context and a more holistic perspective on the world of Gen AI.

- Watch this video to see BigDog in action and to learn more about the technology.

Siri (2010)

In 2010, Apple launched Siri, a speech-recognition application that marked a breakthrough in the field of AI. When Siri was launched in 2010, it had the ability to connect with 42 web services to search for the answer to the question posed. By 2011, Siri was expanded to include multiple languages to ultimately operate internationally (Bosker, 2013).

Explore Deeper (Recommended)

For those who wish to venture beyond the main learning path, this section offers a deeper dive into fascinating sub-topics. These materials are optional, but they provide a richer context and a more holistic perspective on the world of Gen AI.

- Learn more about Siri, Apple's natural language processing software.

Watson (2011)

The IBM-built machine Watson was used to play an American gameshow called Jeopardy!, which proved to be a far greater challenge than the 1997 chess test against Garry Kasparov. IBM made use of neural networks and trained them over a period of three years to enable the machine to recognise patterns in questions and answers that may be used in the gameshow. The victory of Watson in Jeopardy! in 2011 was dubbed "a triumph for AI" (BBC, n.d.).

Explore Deeper (Recommended)

For those who wish to venture beyond the main learning path, this section offers a deeper dive into fascinating sub-topics. These materials are optional, but they provide a richer context and a more holistic perspective on the world of Gen AI.

- Watch Watson play the gameshow Jeoparty! against two opponents.

DeepMind (2012)

In 2010, DeepMind was founded in the United Kingdom. The company's mission was to develop computer programs that can solve complex problems without being taught how to do it (DeepMind, n.d.). Soon thereafter, it was acquired by Google in 2014, a move that seemed fairly odd to the AI community because DeepMind was seemingly irrelevant at the time. The reason for the acquisition, as it became clear at a later stage, was founded in the progress DeepMind had made in machine learning (Wooldridge, 2018:42)..

3. Recent developments

At the heart of the latest, and persisting, surge in interest in AI is the proposal that deep neural networks can learn efficiently from large datasets. The centre of this is "deep learning", which uses large artificial neural networks and applies them to a wide range of tasks. {Neural networks and deep learning are covered in detail later.} In addition to more powerful algorithms, advancements are made at a rapid pace because of the increased access to vast computational power, as well as the access to, and low cost of, storing large amounts of data (Vander Ark, 2018; Parloff, 2016).

It is important to point out some of the significant steps in AI's history behind this recent wave of innovation, which have produced giant leaps in the quality of everyday technologies. As a result, AI has become ubiquitous and pervasive. It is being integrated and embedded in systems and applications in numerous areas of people's lives. For this reason, some believe that AI will transform all industries in the same way electricity has (Burgess, 2018).

The first building block was the work by Yann LeCunn and Geoff Hinton, who, in 2012, applied deep learning to the problem of image recognition. They produced a step change in image recognition accuracy, which led many other researchers to develop increasingly better images, and, soon, speed recognition algorithms as well. Daily interaction with these innovations take place through, among other things, social media, CCTV video processing, and border control, but AI technologies have far greater impact in other sectors, such as medicine, in which they assist in diagnosis and life-saving tasks (Parloff, 2016). One of the most interesting aspects is that, after the initial training of the input algorithms, no further programming is done by humans to enable these machines to perform their tasks. Humans merely import an algorithm into the machine, which then not only completes certain tasks, but also outperforms humans in the process (Parloff, 2016).

The healthcare industry illustrates both the great successes, such as the innovation potential, but also the limitations of AI. One example is using AI to augment diagnostic procedures. For example, in radiology. AI systems have demonstrated their ability to diagnose retina degeneration or pathological lung formations more reliably than their human doctor counterparts. With the need to quickly respond to new medical challenges like the coronavirus (COVID-19) pandemic, AI has proven to be very helpful. At Oxford University, researchers showed how past patient data, coupled with new COVID-19 patient data, could help to predict whether new patients admitted to the emergency department were likely to be suffering from COVID-19, and thus arrange their treatment accordingly.

Explore Deeper (Recommended)

For those who wish to venture beyond the main learning path, this section offers a deeper dive into fascinating sub-topics. These materials are optional, but they provide a richer context and a more holistic perspective on the world of Gen AI.

- Read about how AI is used in diagnostic procedures to read x-rays, and Oxford University's application of AI technology to predict COVID-19-positive patients.

AI has thus been able to harness its great power to recognise patterns in large data sets to make accurate predictions. However, two things are needed here: a clearly defined task and a significant amount of training data to tune the algorithm. Many healthcare operations may clear the data hurdle, but the complexities caused by individual cases that do not fit the general mould often prevent AI from unfolding its true power. However, the goal is to eventually create artificial general intelligence, where a machine will be able to answer any new question without having to be retaught. Nevertheless, this technology is still well beyond the horizon (Wooldridge, 2018). Still, it is this artificial general intelligence that prominent critics, such as Elon Musk, are warning about, because of the vast amount of power it represents and the potential impact it may have in society (Sulleyman, 2017).

Explore Deeper (Recommended)

For those who wish to venture beyond the main learning path, this section offers a deeper dive into fascinating sub-topics. These materials are optional, but they provide a richer context and a more holistic perspective on the world of Gen AI.

- Read the following review to learn more about the complexities of individual healthcare cases, and how IBM Watson overpromised and underdelivered using AI.

The second innovation was the application of neural networks to reinforcement learning by Ian Goodfellow in 2014. This is covered in more detail {later}, but this machine learning approach enabled image and video technologies, including image generation and the infamous deepfakes. It was also an important building block for generative AI, which includes natural language processing applications like ChatGPT or Claude. The third innovation was the development of so-called “transformer” architectures, which were introduced in 2017 by Google researchers.

Taken together, the ideas of deep learning, reinforcement learning, and transformers have driven the staggering speed at which AI is developing. Innovations like ChatGPT, for example, would not have been possible without them. GPT, which stands for “Generative Pre-Trained Transformer”, directly applies these ideas to understanding and generating text. Yet the same technology can be used on any kind of data, from text, to images, to videos, to sound. Thus, we can expect this current wave of innovation to produce many more innovations, beyond text generators like ChatGPT, or image generators like Dall-E2, Imagen, or Stable Diffusion, that have captured the attention of a global audience.

Alongside these drastic innovations that have taken place since 2010, a key emerging

concern when it comes to AI applications is the future of work, or more specifically, the displacement of jobs. Here, it is not the case that AI will render large parts of the white- collar workforce redundant, but rather that it will augment their abilities through the tools it creates. Hence, as the price for products and services fall, AI may even hold the potential to create a net increase in employment due to its contribution to productivity (Vander Ark, 2018). In {Module 4}, you will learn more about the future of work and how organisations can best prepare their workforces to use machines to their advantage.

However, another major concern relating to AI is the potential that it will become smarter than humans. The late Stephen Hawking also warned that AI could potentially lead to the end of the human race (Cellan-Jones, 2014). Regulation of AI technology is key to contain the threats of AI (Bedford, 2018). The legal and regulatory considerations around AI will be explored in more detail in Module 5.

Despite the warnings issued relating to the advancement in AI technology, not all AI researchers agree with the “doom and gloom” attitude. as alluded to by the late Stephen Hawking. Google CEO, Sundar Pichai, is of the opinion that AI must be pursued at an increasing rate due to the potential that it holds to increase the lifespan of humans, through, for example, improving medical diagnosis of cancer (Clifford, 2018).

Explore Deeper (Recommended)

For those who wish to venture beyond the main learning path, this section offers a deeper dive into fascinating sub-topics. These materials are optional, but they provide a richer context and a more holistic perspective on the world of Gen AI.

- Consult the following source for more information on the current landscape of AI and what's new.

Appendices

Appendix 1: Reference list

AI Programme (2025). Course notes. University of Oxford, Saïd Business School.

Bloggs, J. (2013). 'The key issue'. [Cartoon]. The Times, 20 January, p.56.

British Standards Institution (2008). BS 8498:2008: British Standards for screw threads and screw gauges : introduction and guidance. London: British Standards Institution. Available at: https://bsol.bsigroup.com/Bibliographic/BibliographicInfoData/000000000030165215. (Accessed: 26 March 2021).

Bureau van Dijk (2013). 'BSkyB plc company report'. Available at: http://fame.bvdep.com/ (Accessed: 8 January 2013).

Custer, C. M. (2018). 'Annual Variation Birds as indicators of contaminants in the Great Lakes'. doi.org/10.5066/F7SB452P.