Artificial Intelligence

Artificial Intelligence:

Benefits and Costs

By

Nicola Acocella

Cambridge Scholars Publishing

Artificial Intelligence: Benefits and Costs

By Nicola Acocella

This book first published 2024

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-0364-1295-1
ISBN (Ebook): 978-1-0364-1296-8

Table of Contents

LIST OF FIGURES

1

DEFINITIONS AND TYPES

Artificial Intelligence (AI) is the computer science field creating machines that mimick human intelligence and use algorithms to simulate human-like behavior and decision-making. 'In AI, machines perform tasks like speech recognition, problem-solving, learning, etc. Machines can work and act like a human if they have enough information'. Then, in AI knowledge engineering plays a vital role by establishing the relationship between objects and their properties (Pedamkar, 2023).

As said by Wodecki (2018), AI covers an interdisciplinary field, integrating notions of logic, mathematics, computer science, psychology and, more recently, management and ethics. It can be developed dynamically as a method and from a technological point of view. It raises many hopes and many fears and controversies (see, e.g., Bostrom, 2014).

In the last 60 years, AI comes after at least four digital revolutions: the advent and proliferation of personal computers; the expansion of the internet and search; the rise and influence of social media; and the growing mobile computing and connectivity.

Some existing definitions of AI, used by G20 economies in the introductory part of their AI modules, are provided below.

According to Eurostat, AI refers to systems using technologies such as: text mining, computer vision, speech recognition, natural language generation, machine learning, deep learning for gathering and/or using data to predict, recommend or decide the best action to achieve specific goals.

AI systems can be:

- *purely software based*, such as chatbots and business virtual assistants, which use: natural language processing; face recognition systems, with the help of computer vision or speech recognition systems; machine translation software; data analysis based on machine learning, etc.;

- *embedded in devices*, as autonomous robots for warehouse automation or production assembly works;

- *autonomous drones* for production surveillance or parcel handling, etc..

Statistics Canada (2019) says that AI refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals. AI-based systems can be purely software-based or embedded in a device.

According to INSEE (2019), AI includes all the technologies aiming at computerisation of cognitive tasks traditionally performed by humans, voice recognition, biometrics, image recognition, decision support, etc. (Istat, 2021).

AI is thus a branch of computer science studying the development of hardware and software systems having specific capabilities typical of the human being (interaction with the environment, learning and adaptation, reasoning and planning), capable of pursuing autonomously a defined purpose, making decisions that until then were usually entrusted to people. In the words of Monostori (2014), "AI is the science and engineering of making intelligent machines, especially intelligent computer programs". AI systems show some traits of human intelligence such as learning, problem-solving, perception, and some creativity and social intelligence (Diaz, 2023).

AI is a research field that studies the programming and design of systems aimed at providing machines with one or more characteristics that are considered typically human. Properties range from learning to visual or space-time perception. AI is a complex topic with many layers.

A 'true' AI machine can simulate human intelligence, behaviour and emotions. While no machine has still reached that level, modern AI can perform moderately complex functions. It can solve problems and decide using data inputs. In addition, it can recognise and interpret visual information as well as respond to written and verbal language. It can do so by using an enormous amount of information, thus being able to process new data, drawing unique, intelligent conclusions based on this information (https://www.frogs-company.com/ai-applications). The information it uses is entirely based on the sources supplied by humans. It is transformed into data and processed in ways that are often uncertain and can be often forecast with difficulty.

After the first studies of the Fifties, the expectations about AI began to increase, but, due to a lack of adequate computational devices, the concept of AI soon fragmented into distinct theories, which are still shared today and identify the three types of AI:

1. *Strong AI*, according to which machines are capable of developing self-awareness, which studies systems capable of replicating human intelligence. This paradigm is supported by the research field of *General AI (AGI)*, which studies systems capable of replicating human intelligence.

2. *Artificial Narrow Intelligence* (*ANI*), also known as *Narrow AI* or *Weak AI*, which believes it possible to develop machines capable of solving specific problems without being aware of the activities performed. The objective of this type is therefore not to create machines equipped with human intelligence, but to have systems capable of performing one or more complex human functions, such as voice assistants of the kind of *Siri*, *Alexa*, *Google assistants, ChatGPT*, which can perform a specific task, such as generating text responses to directions received.

3. *Artificial Super Intelligence* (*ASI*) - or *Super AI* -, possibly to be considered as a way into the future, which will be achieved when AI is more capable than a human.

Let us gain a better understanding of each type. 1. AGI aims to build machines that think and would have the ultimate ambition to produce a machine whose overall intellectual ability is indistinguishable from that of a human being; it would represent generalised human cognitive abilities in software; therefore, even when faced with an unfamiliar task, the AGI system could find a solution.

AGI would be an intelligent system for computers and have comprehensive or complete knowledge and cognitive computing capabilities. This system would be indistinguishable from that of a human, except for the incredible high speed that today no existing computer can achieve. Such a system should be creative, have sensory perception, recognise colour, depth and three dimensions in static images, pinpoint a geographic location, etc. However, it can also cause dangers, as Stephen Hawking warned in 2014: "The development of full AI could spell the end of the human race. It would take off on its own and redesign itself at an ever-increasing rate. Humans, who are limited by slow biological evolution, couldn't compete and would be superseded."

2. In contrast to AGI, ANI is in practical use today and is able to perform specific tasks or types of problems, by using a combination of machine learning, deep learning, reinforcement learning and natural language processing (see below) for solving specific types of problems and executing specific focused tasks. It is goal-oriented without the ability of a self-expanding mechanism.

Examples of this type of AI include: customer service chatbots; voice assistants of the kind of Apple's Siri and Amazon's Alexa; recommendation engines (e.g., Google, Netflix, Spotify); self-driving cars, recognising other vehicles, people and objects while adhering to driving rules and regulations; AlphaGo, a computer program that excels in a specific type of problem solving.

There are different types of narrow AI: reactive AI, which is the basic version, with no memory or data storage capabilities, emulating the human mind's behaviour and responding to different interpretations without any prior experience; limited Memory AI**,** more advanced, having great memory and data storage capabilities enabling machines to offer more precise interpretation using statistical data. Examples of narrow AI are: virtual assistants (e.g., Siri, Alexa, Cortana); IBM's Watson, which is capable of answering questions asked in natural language interpretations; Google's Rankbrain, an algorithm used to sort the search results; self-driven cars; facial/image recognition; drone-based robots.

Narrow AI has limitations due to its limited capabilities and the possibility of causing harm in case of a system failure, such as for self-driving cars, if they miscalculate the location or have a design flaw.

Another limitation might lie in the loss of jobs caused by narrow AI, but such loss might be countered by an increase in the job opportunities in other fields.

3. Artificial superintelligence would be capable of surpassing human intelligence by manifesting cognitive skills and developing emotional understanding, beliefs, skills and desires of its own. It would be the most advanced, powerful, and intelligent type of AI, transcending every human being. Superintelligent machines would be self-aware and could have all the tasks humans can as well as non-human abstractions and interpretations.

Super AI Systems could imply threats, leading to a global catastrophe, due to loss of control and understanding. In fact, ASI systems could use their power and capabilities to carry out unforeseen actions and eventually become unstoppable, without the human possibility to contain them once they emerge or as they become the weapon of a government, a dictator, or social class or a corporation, and even sociopaths to unleash nuclear attacks, autonomous weapons, drones, and robots.

Super AI has, however, also potential advantages in reducing human error, being helpful in giving access to millions of programs. This can be done by: building logic from the available data on its own, compiling and debuging programs; replacing humans to accomplish risky tasks such as defusing a bomb, exploring the deepest parts of the oceans, coal and oil mining, dealing with the consequences of natural or human-induced disasters; being always available; and for the possibility to explore new science frontiers (as for space travels).

At the time of the Chernobyl nuclear disaster, in 1986, superintelligent robots were not yet invented. If they were, they could intervene and eliminate the nuclear power plant's radiations, which were so intense as to kill any human approaching the place. Authorities were forced to pour sand and boron from a distance above by using helicopters. In addition, the healthcare industry could benefit significantly from super AI, which can play a pivotal role in drug discovery, vaccine development, and drug delivery, e.g., for intracellular drug delivery.

Another advantage of super AI is that machines can be programmed to work 24 hours per 7 days a week, without any breaks, in contrast to human interventions. This could be done in a number of occupations from educational institutes to helpline centers.

Finally, space exploration could be facilitated by super AI, by testing and estimating the probability of success of many equations, theories, researches, rocket launches, and space missions.

In this scenario, AI must be treated by combining the theoretical aspects with the practical and operational ones.

Starting from the meaning of AI, we can here describe the main AI techniques (*Machine Learning* and *Deep Learning* above all). Examples of machine learning are speech and image recognition (e.g., when social media network can recognise faces), and protection from fraud. This is only possible because a system has been trained on large amounts of data, thus being able to learn from mistakes and devise predictions and decisions.

After analysing the various applications of AI, a question arises: how does AI work? There are different methodologies applied to this research field.

In fact, the applications and history of AI intersect with that of Machine Learning. If it is true that AI is configured as the study of the development of systems endowed with capabilities typical of human beings, Machine Learning can be considered a (very popular) path for the application of AI. We therefore refer to systems capable of learning from experience, with a mechanism similar to what a human being does from birth.

Based on the learning techniques, it is then possible to distinguish different types of Machine Learning, which can translate into different applications. The best known of these is certainly *Deep Learning*, but Machine Learning is much more multifaceted. In any case, by using *Deep Learning* technology, complex patterns can be recognised. We look here at its six main methodologies.

- *Model Prediction* or *Predictive Modelling*, is an area of Machine Learning that includes a variety of techniques capable of gathering information from available data and learning models to apply to new data, which can be done by these models. Predictive Modelling therefore aims to make predictions about data or events in the future. This AI application channel is becoming increasingly popular in a number of companies and is especially applied in the field of fraud prevention, predictive maintenance and qualitative analysis.

- *Deep learning* is one of the most important areas of machine learning. It is a set of techniques that simulate the learning processes of the brain through layered artificial neural networks, in which each layer calculates the values for the next one. Simply put, Deep Learning is a learning technique in which artificial neural networks are exposed to vast amounts of data; these are then able to learn by themselves to perform certain tasks without the need for data pre-processing, such as speech and image recognition.

- *Online Learning*, also known as *Real Time Machine Learning*, is the branch of Machine Learning that studies techniques for problems in which data become available sequentially, one after another, and decisions must be made as a piece of data becomes available. Unlike *Offline Learning*, whose decisions are based on a previous acquisition of a set of data, learning and decision making are interrelated and interdependent: what is learned conditions the decisions, which condition the data that is observed for further learning.

- *Explainable Regression and Classification* consists of Machine Learning techniques aimed at solving regression problems (prediction of ordinal quantities) and classification, such as to allow the explanation of the results obtained. In other words, Explainable Regression and Classification allows one to understand why AI models for regression and classification make certain decisions following data processing, making this person to understand what happens in the 'black box' of Machine Learning.

- *Information Retrieval (IR).* Information Retrieval is the set of techniques used to manage the representation, storage, organisation and access to objects containing information such as documents, web pages, online catalogs and multimedia objects. The ultimate goal of these processes is to provide the user with information that he has previously searched for and which is, therefore, more relevant to his interests.

- *Reinforcement Learning.* Reinforcement learning is a machine learning technique aimed at solving sequential decision-making problems by achieving certain objectives through interaction with the environment in which it operates. An agent learning with reinforcement algorithms learns to make optimal decisions through an empirical trial-and-error approach. At each iteration the agent observes the environment through its sensors (physical or virtual), decides which action to perform and observes the effects that the chosen action has had on the environment[1].

[1] Readers of Chen, Chen (2022) are guided through the history and development of AI, beginning from its early start and going to its potential future applications.
A general, historical and methodological analysis is in Russell, Norvig (2020).

2

SPACE-TIME HISTORY

In this chapter I hope to provide a comprehensive history of AI right from its lesser-known days (when it wasn't even called AI) to the current age of Generative AI.

2.1. Myths, dreams and AI

Gold (2023) rightly speaks of the perennial interest of humans in making machines that displayed intelligence: ancient Egyptians and Romans were attracted by religious statues manipulated by priests behind the scenes. The Greek myths of Hephaestus and Pygmalion were based on the idea of intelligent automata (e.g., Talos, the bronze giant forged by Hephaestus presented to Minos and guardian of the Island of Crete) and artificial beings (e.g., Galatea and Pandora).

Automation concerns are not new to AI. Other examples of similar concerns date back even to the advent of written language. In fact, in the IV century B. C., Plato's Phaedrus described displacement of human memory and reading by writing. More recently, the Industrial Revolution and the Luddite riots of 19th-century are examples of technological advancement leading to social unrest.

The preoccupation with developing practical methods for making machines behave as if they were humans emerged already 7 centuries ago.

In fact, this interest was also common to medieval people. In 1308 the Catalan poet and theologian Ramon Llull perfected his method, founded on the use of paper-based mechanical means for creating new knowledge from combinations of concepts (see his *Ars generalis ultima*, i.e. The Ultimate General Art). The 16th century Swiss Philosopher Theophrastus Bombastus said, “We shall be like gods. We shall duplicate God's greatest miracle – the creation of man."[1] Llull's proposal was followed by the mathematician and philosopher Gottfried Leibniz, who published *Dissertatio de arte combinatoria* (On the Combinatorial Art), where he argued that all ideas derive from combinations of a relatively small number of simple concepts. The modern Swiss alchemist Paracelsus aimed at fabricating an artificial man by human fecundation of a horse. There are traces of an alchemically-fabricated immortal homunculus also in Goethe's Faust. Instances of AI are also in fiction, such as Frankenstein's. But there are also examples of devices anticipating modern applications of AI, such as Al-Jazari's invention of a method for controlling the speed of rotation of a wheel by using an escapement mechanism[2]. Also the novelist Jonathan Swift in his 1726 Gulliver's Travels described the 'Engine', a machine for improving speculative knowledge by practical and mechanical operations.

2.2. Contemporary attempts and realisations

As said, our species' latest attempt at creating synthetic intelligence is now known as AI.

The period between 1940 and 1960 was strongly marked by the conjunction of technological developments (of which the Second World War was an accelerator) and the desire to understand how to bring together the functioning of machines and organic beings. For Norbert Wiener, a pioneer in cybernetics, the aim was to unify mathematical theory, electronics and automation as 'a whole theory of control and communication, both in animals and machines'. As early as 1943, a first mathematical and computer model of the biological neuron (formal neuron) had been developed by Warren McCulloch and Walter Pitts. The two researchers thus formalised the architecture of our contemporary computers and demonstrated that they were universal machines, capable of executing what is programmed.

Turing, who was a British mathematician, on the other hand, raised the question of the possible intelligence of a machine for the first time in his famous 1950 article, 'Computing Machinery and Intelligence' (see Turing, 1950), and described a 'game of imitation', where a human should be able to distinguish in a teletype dialogue whether he is talking to a man or to a machine (see more below). However controversial this article may be (this

[1] For a more complete indication of these ancient interests in AI see Wikipedia, Timeline of AI.
[2] Al-Jazari was an Arab mathematician, inventor and mechanic engineer living in the 12th century.

'Turing test' – the term originally indicating the imitation game - does not appear to qualify for many experts), it will often be cited as being at the source of the questioning of the boundary between the human and the machine. Some aspects of the Turing test were prefigured by René Descartes in his 1637 *Discours de la Méthod*, when he imagined a machine being constituted so that it could utter words, and even emit some responses to action on it.

The Dartmouth conference - held at Hanover, New Hampshire (USA) in 1956 - was the event that in contemporary times 'initiated AI as a research discipline', which grew to encompass multiple approaches, from the symbolic AI of the 1950s and 1960s to the statistical analysis and machine learning of the 1970s and 1980s and then to today's deep learning.

The term AI was first coined by John McCarthy in the seminal event in the history of AI just mentioned, the Dartmouth Conference, which had a significant impact on the overall history of AI, by helping to establish AI as a field of study and an academic discipline, encouraging the development of new technologies and techniques. In addition to McCarthy, the participants included Marvin Minsky, an American scientist, co-founder of the MIT's AI laboratory, and Herbert Simon, an American political scientist, who influenced computer science, economics, and cognitive psychology. Together with other prominent scientists and researchers, they discussed a wide range of topics related to AI and laid out a roadmap for AI research, which included programming languages and algorithms for creating intelligent machines (natural language processing, problem-solving, and machine learning).

Even if the term AI was first coined – as said - by McCarthy at the beginning of 1950, John Von Neumann - the Hungarian-born mathematician - author, together with Oskar Morgenstern, of the game theory - and Alan Turing were the founding fathers of the technology behind it, by making the transition from the 19th century decimal logic (dealing with values from 0 to 9) and machines to binary logic (relying on Boolean algebra and dealing with chains of 0 or 1). In particular, Alan Turing can be rightly considered one of the fathers of modern computer science and of AI. In his article 'Computing machinery and intelligence', he proposed the test (i.e., the *Turing Test*) for defining a machine as intelligent. According to this test – which remains a benchmark for measuring the progress of AI - a machine could be considered intelligent if its behaviour was indistinguishable from that of a person[3].

The Dartmouth conference was the first of its kind, in the sense that it brought together researchers from seemingly disparate fields of study – Computer Science, Mathematics, Physics, and others – with the sole aim of exploring the potential of Synthetic Intelligence.

In 1958 the psychologist Frank Rosenblatt designed an artificial neural network architecture, called the *Perceptron*, which is a binary classifier of input patterns into 2 categories. After a period of decline, the Perceptron was incorporated into more complex networks, contributing to the development of deep learning.

The 1960s were a period of significant progress in the development of AI, when computer scientists and researchers explored new avenues for creating intelligent machines and corrected some flaws of the machine learning algorithms previously suggested, thus leading researchers to explore other AI approaches, such as symbolic reasoning, natural language processing, and machine learning, and also leading to the development of new programming languages and tools (the *General Problem Solver* - *GPS* - created by Herbert Simon and others, which could solve problems by searching in a space of solutions; and the *ELIZA program*, created by Joseph Weizenbaum, a natural language processing program that simulated a psychotherapist).

The 1980s, instead, were a period of significant slowdown of research and development in the field of AI, due to the failure of many of the programs developed in the 1960s and the following years, thus leading to a significant decline in funding for AI research – which had been copious before, in

[3] The importance of computation for human bodies was stressed by the 17th century philosophers Leibniz, Hobbes and Descartes. These three authors analysed the possibility of making all rational thought as systematic as algebra or geometry. In fact, Hobbes' Leviathan says that 'reason is nothing but reckoning'. Leibniz explored the route for finding a universal language of reasoning reducing argumentation to calculation, in a way to eliminate disputations between philosophers. French philosopher and mathematician Descartes in his *Discours de la Méthod* already mentioned was interested in automata, which, he said, can't respond to things the way a human can. His analysis served as a basis for the Turing test.

particular in the United States and United Kingdom - and an understandable stagnation in the progress of AI.

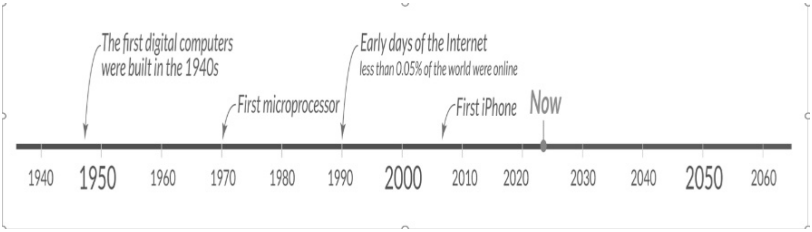
Nevertheless, an important type of AI – the *Expert Systems* - was developed in the 1980s. This mimics the decision-making abilities of a person that is expert in a specific field, such as medicine (diagnosing medical conditions), and finance (predicting stock prices). In the latter case, this provided significant benefits to businesses and thus served as a proof that AI systems could be used in real life. Expert systems have been significant in the history of AI and continue to be used in various industries, being at the root of machine learning and natural language processing. However, they are also criticised, e.g., for requiring a high amount of different information and having a superficial knowledge, possibly choosing the most inappropriate method for solving a particular problem and being potentially expensive from the point of view of computation. By contrast, a form of neural network, called 'Hopfield net', was introduced that was able to learn and process information in a new way. An application to optical and speech recognition was instead offered in 1986. Creation of Robotics was derived from cybernetics and control theory, starting from the late 1980s.

AI was in its worse days when in 1989 Carnegie Mellon – and later IBM - began work to create *Deep Thought*, which was later renamed *Deep Blue* – the AI machine that defeated the chess game champion Garry Kasparov the second time they met in 1997. The forecasts of the pioneers of the 1950s had failed, due to the attempt to use pure logic by the scientists applying AI. That ended with *Deep Thought* and *Deep Blue*. Afterwards, a radically more promising form of AI, the neural net, was developed. This, however, required lightning-fast computers, tons of memory, and lots of data, not readily available at the time. Even later, in the 1990s, neural nets were considered a waste of time. Only in the 2000s, computers evolved in a way to make neural nets viable. A huge industry in ultrafast graphic-processing units derived from the passion of video-game players for ever-better graphics. This turned out to be perfectly suited for neural-net math. And the explosion of the internet produced a torrent of pictures and texts that could be used to train the systems.

By the early 2010s, Geoffrey Hinton – named the 'godfather' of AI - and his crew of 'true believers' could take neural nets to new heights. Networks with many layers of neurons could be created (materialising the 'deep' in 'deep learning'). In 2012 they won the annual *Imagenet* competition, where AI competes for recognising large-scale objects. Then self-learning machines were finally viable.

Figure 2.1. gives a summary look at some of the steps in the progress of AI, from the first digital computers to internet and the iPhone.

Figure 2.1. Steps in the progress of AI, Source: Roser (2022)



In the 1990s – years representing a major milestone in the history of AI - the *Natural language processing* (NLP) had been developed, combining computational linguistics and rule-based modelling of human language, together with Computer Vision systems, allowing for more accurate object recognition and image classification. These techniques are now used in a wide range of applications, from self-driving cars – such as that built at Stanford University, which drove for 131 miles in the desert in 2005, or the one navigating in 2007 for 55 miles in a town - to medical imaging (see Kumar et al., 2023).

In the early 2000s research on *big data* - i.e., data being of a high volume, of a high variety and quickly-generated – rose to prominence, providing massive amounts of data, derived from various sources, including media, sensors, etc. At the same time, rising capacity of data storage and of data processing allowed elaboration and analysis of large datasets as well as accurate predictions.

The emergence – in the course of many decades - of *Deep Learning* is a milestone in modern AI. After the appearance of big data, existence only of manually programming machines made it difficult to use them for certain applications - being limited in their ability to learn and adapt to new data - and researchers needed new ways to process and derive insights from vast amounts of information. This was made much easier by deep learning algorithms, using artificial neural networks[4] – "made up of layers of

[4] The term 'neural network algorithm' indicates a collection of AI algorithms that mimic the functions of a human brain. These tend to be more complex than many other algorithms and have applications beyond some of the ones discussed here. In

interconnected nodes, each of which performs a specific mathematical function on the input data" (Gold, 2023) - modelled on the structure and function of the human brain. These algorithms provided a solution by enabling machines to learn from large datasets and make predictions or decisions. In these networks the output of one layer serves as an input to the next, allowing the network to extract increasingly complex features from the data.

One of the key advantages of deep learning is its ability to learn hierarchical representations of data. This means that the network can automatically learn to recognise patterns and features at different levels of abstraction. For example, a deep learning network might learn to recognise the shapes of individual letters, then the structure of words, and finally the meaning of sentences, thus possibly generating realistic human-like language.

Generative AI is the latest development - and a major milestone - in AI. It involves creating AI systems capable of generating data that are new or similar to data they were trained on, such as images, text, music, and videos. It has been revolutionised by transformers, which are a type of neural network architecture that uses self-attention mechanisms to analyse the relationships between different elements in a sequence. This allows them to obtain more consistent and nuanced output, developing large language models such as *GPT-4 (ChatGPT)* - which can generate a human-like text on a wide range of topics - as well as AI art[5]. This text "can be created by new and unique pieces of art in the field of creative writing with some authors using them to generate new text or as a tool for inspiration" (Gold, 2023) or even novels, codes, poetry, and other reading and writing tasks. This can be done by elaborating information available in its capable memory and assembling this information, as forecast in 2004 by Philippe Vasset – a French novelist – in his *ScriptGenerator*[6]. The Chinese Alibaba has created a language-processing AI, beating human intellect for a Stanford reading and comprehension test. A humanoid robot named *Sophia* has been created, being the first robot having a realistic human appearance, able to see, to communicate, and to replicate emotions. Before that, the Japanese Waseda University had initiated a project that was completed in 1972,

---

unsupervised and supervised learning, it can be used for classification and pattern recognition

[5] See also Korinek (2023).

[6] See Vasset (2005).

building the *WABOT-1*, which was the first humanoid robot, able to walk and transport objects.

Some of these developments of AI have generated a large debate on the ethics of its use[7], which has added to the critiques by philosophers about definitions and internal consistency of the AI program[8].

As Roser (2022) notes, "computers and AI have changed our world immensely, but we are still at the early stages of this history. … AI has already changed what we see, what we know, and what we do. … There are no signs that these trends are hitting any limits anytime soon. To the contrary, particularly over the course of the last decade, the fundamental trends have accelerated"[9].

Haenlein, Kaplan (2019) offer a comprehensive outlook on the past of AI and its future perspectives, drawing on aspects that are micro (for regulation with respect to algorithms and organisations), meso (as to regulation with respect to employment), and macro (as to regulation with respect to democracy and peace). The capabilities of AI systems for language and image recognition have developed very rapidly and are now comparable to those of humans, considering that 10 years ago, "no machine could reliably provide language or image recognition at a human level. … AI systems have become steadily more capable and are now beating humans in tests in all these domains".

The time is ripe for Generative AI becoming useful for the regular, non-tech person. GPT-4 and Gemini, are multimodal, as they can process not only text but images and videos, thus being capable of unlocking many new apps. 'For example, a real estate agent can upload text from previous listings, fine-tune a powerful model to generate similar text with just a click of a button, upload videos and photos of new listings, and simply ask the customized AI to generate a description of the property.'

The success of this plan, however, depends on whether these models work reliably, which is not always the case, e. g., language models and generative models are riddled with biases. In addition, they can be hacked, especially if they are allowed to browse the web (Heikkilä, Heaven, 2024).

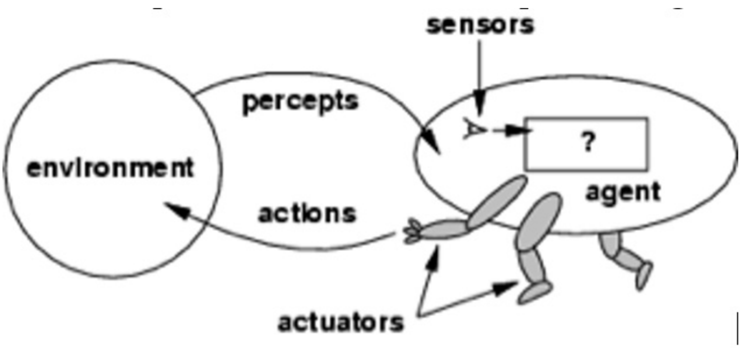
[7] An example of these ethical critiques is offered by Weitzenbaum (1976).
[8] See, e.g., Lucas (1961).
[9] In a few years, 90% of internet messages could be produced by AI.

2.3. The way a contemporary intelligent agent operates

Perceiving human environment and taking actions maximising chances of success is the object of intelligent agents *as a system*. The *paradigm of intelligent agents* allows to study isolated problems and find solutions that are both verifiable and useful by drawing on concepts from decision theory and economics.

Russell, Norvig (2009) deal with the intelligent agent who perceives from the environment and acts, such as in the following figure 2.2.

Figure 2.2. An overview of the actions of an intelligent agent, Source: Stanford Encyclopedia of Philosophy (2018)



Probabilistic reasoning applied to AI added rigor, because almost every task requires that conclusions are drawn from uncertain clues and incomplete information. It does so by reasoning with partial belief, being useful in a number of fields, such as diagnosis, forecasting, multi-sensor fusion, decision support systems, plan recognition, and planning (Russell, Norvig, 2021).

Outside these standardised tests the performance of the mentioned types of AI is mixed. In some real-world cases these systems are still performing much worse than humans. On the other hand, some implementations of such AI systems are already so cheap that they are available on the phone in your pocket: image recognition categorises your photos and speech recognition transcribes what you dictate.

An image summary of AI developments through time and a comparison of the relative abilities of humans and AI is offered by figure 2.3.

Figure 2.3. Steps of language and image recognition capabilities of AI systems, Source: Roser (2022)

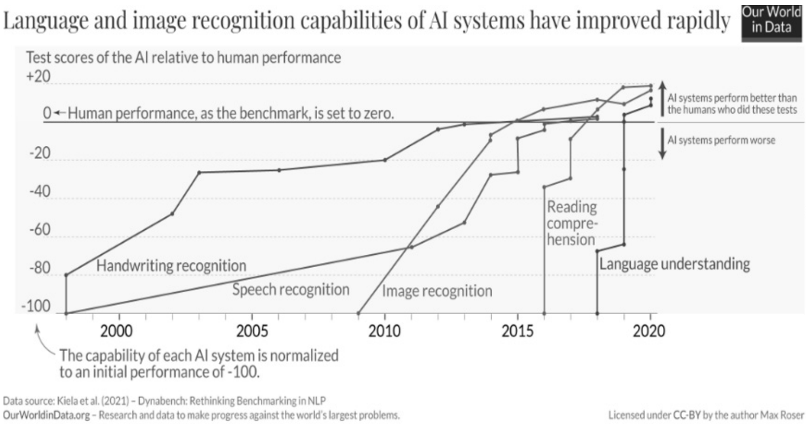
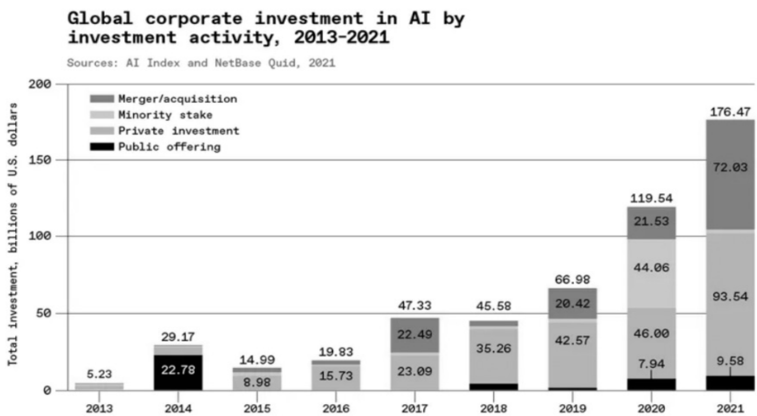


Figure 2.4 shows the huge rise in 2020 and 2021 of corporate investment in AI by type of investment, especially through mergers and acquisitions and private investment.

Figure 2.4. Corporate investment in AI by investment activity, 2013-2021, Source: Strickland (2022)



Note: For all years, except the first two, investment in merger and acquisition is on the top, with the other types below in the order indicated in the caption, whereas investment in minority stakes are on the top, with private investment below for both 2013 and 2014 and public offerings even below in 2014.

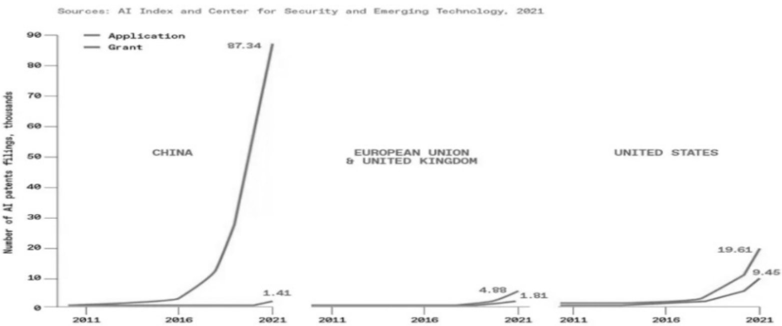
Image generation has been the step of AI development following image recognition. In fact, AI systems are now able to generate or counterfeit images of faces.

The current engagement is now in language recognition and production, creating systems that analyse human language and respond to it. One of such devices is *PaLM*, a system developed by Google. But there are also: auto-completion of e-mails, translation of texts, automatic transcription of videos.

AI systems are not yet able to produce long, coherent texts or summarize them. In the future, we will see whether the recent developments will slow down – or even end – or whether we will one day read a bestselling novel written by AI. AI-based models are forecast to cut greenhouse gas emissions by 4% within the year 2030.

Figure 2.5 offers an idea of applications and grants for AI in various countries. In 2021, in contrast with the huge number of China's applications (52% of world total), it was the United States that dominated for grants (40% of the global total). A grant certifies that the patent is actually credible and useful.

Figure 2.5. Applications and grants for AI in various countries, 2011-2021, Source: Strickland (2022)



3

MEASUREMENT

3.1. The various ways for measuring developments related to AI

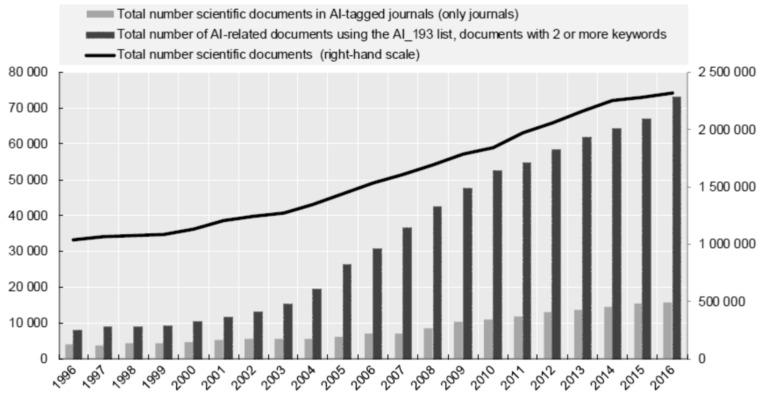
Measuring AI is difficult, but not impossible.

In 2018, the OECD initiated work aimed at identifying and measuring developments related to AI, through data on scientific publications, on open-source software, and on patent filings.

In parallel to the OECD efforts, in 2018 and 2019, several institutions and research groups – such as the China Institute for Science and Technology Policy at Tsinghua University, the Stanford University Joint Research Centre and the European Commission - proposed alternative approaches to measure AI by using data on scientific publications and/or on patents.

The first way to measure developments of AI is related to scientific documents in AI-tagged journals (see figure 3.1 and 3.2).

Figure 3.1. Total number of scientific documents, documents in AI-tagged journals (ASJC) and AI-related documents using the AI_193 keyword list, 1996-2016, Source: Baruffaldi et al., 2020

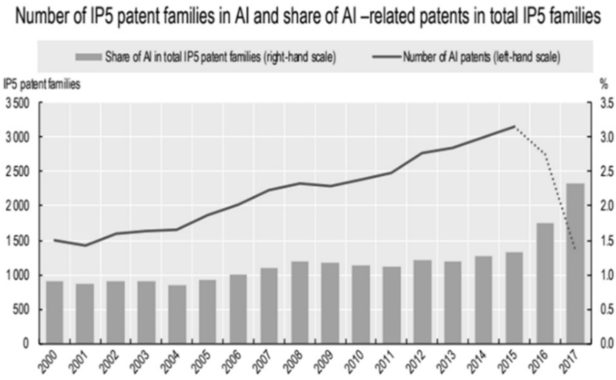


Note: AI-tagged journals only include documents in journals. There is a possible year misclassification issue for conference proceedings in Scopus® which would require further investigation.
Source: OECD calculations based on Scopus Custom Data, Elsevier, Version 1.2018.

The second way to measure AI developments is with reference to open-source repositories related to AI. Since 2014, the number of these repositories has grown about three times as much as the rest of open-source software (Baruffaldi et al., 2020).

The third way of measurement is offered by patents related to AI. Figure 3.2 below gives an account of the development of this source between 2000 and 2017.

Figure 3.2. Trends in AI-related patents, 2000-17, 1996-2018, Source: Baruffaldi et al., 2020[1]



Note: Counts of IP5 patent families are reported according to the earliest filing date of patents that belong to the same family. Data from 2016 are truncated due to unpublished patent data.
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, July 2019.

Another way of measuring the relevance of AI is offered by the revenues derived from it. Revenues from the software market are expected to reach 126 billion dollars worldwide by 2025. Other ways are offered by the institutions that have implemented AI, which are 37% of all organisations. Furthermore, the enterprises employing AI grew 270% over the past four years. Finally, by 2025 95% of customer interactions will be powered by AI and the global AI software market is expected to grow approximately 54% year-on-year, reaching a value of 22.6 billion dollars (Biswal, 2023).

According to Chollet (2019), the need arises of defining and evaluating AI in a way that enables both comparisons between two systems and comparisons with humans. There has been an abundance of attempts to define and measure AI, by comparing the skill exhibited by AI and humans for specific tasks. A new formal definition of Intelligence is based on Algorithmic Information Theory. It describes Intelligence as skill-acquisition efficiency, highligting the concepts of scope, difficulty of generalisation, priors, and experience. This is the *Abstraction and Reasoning Corpus* (ARC), which can be used to measure a human-like form

[1] Figure 3.2 speaks of IP5. This term denotes inventions protected in at least two jurisdictions, at least one of which needs being one of the five patent classes of the PatentsView platform.

of general fluid intelligence, enabling fair general intelligence comparisons between AI systems and humans.

On the policy side, it is crucial to assess the impact of AI on technical progress and the society. Mishra et al. (2020: 2) identify various challenges deriving from this task, such as: "How can we measure the economic impact of AI, especially labor market dynamics but also interaction with economic growth and well-being? … How can we measure the societal impact of AI, particularly on sustainable economic development and the potential risks of AI to diversity, human rights, and security? … How can we measure the risks and threats of deployed AI systems?"

Ochella, Shafiee (2021) deal with the use of AI algorithms in the field of prognostics and health management of systems, sub-systems or components along their lifecycle. In particular, they help in predicting the remaining useful life of mechanical systems, for making informed decisions regarding maintenance and life-cycle planning of assets.

Welty et al. (2019) introduce metrology as a science of measurement of AI systems. They also apply it to human powered evaluations and evaluate the performance of AI systems. They show that metrology can be a guide for turning benchmark datasets into instruments and can improve the science of measuring AI system. Finally, they evaluate AI systems interpreting unstructured inputs, such as images and videos, by using benchmark standard datasets.

3.2. Recent advances in measurement

Van der Maas et al. (2021) look at recent advances in AI represented by deep learning and reinforcement learning. Such advances have deep roots in psychology, in particular for the study of individual differences and cognitive development. Fruitful interactions for the measurement of natural intelligence and AI are expected.

Avrin (2021) proposes a taxonomy of AI capabilities and generalises it to various AI tasks for drawing a parallel with human capabilities and developing a methodology for comparison between the two. It also recommends further actions for identifying the strengths and weaknesses of AI vs. human intelligence.