# Chapter 3 - A brief history of Artificial Intelligence

## From Looms and Mechanical Turks to AlphaZero

### 3.1 - Mechanical Turk (1770)

The Mechanical Turk, also known as the Automaton Chess, was a device that seemed capable of playing chess against a human opponent. Constructed in the late 18th century, it played chess at a remarkable level, but it was all an elaborate hoax. The illusion was created by a human chess master hiding inside the device. Even so, it still stands as a colorful example of people's interest in the possibility of machines showing intelligent behavior.

|  |  |  |
| --- | --- | --- |
|  | A picture containing text, book  Description automatically generated |  |
|  | **Figure 3.1** - The trickery behind the Mechanical Turk |  |

### 3.2 - Luigi Galvani (1780)

Luigi Galvani is famous for his work in electrophysiology, or the study of the nature of electrical impulses generated by our nervous system. Working with dissected frog legs, he noticed that applying current to them made them move as if the animal was still alive. These observations led him to conclude that animals and human bodies contain an electrical fluid, which he called 'animal electricity.' Although it turned out not to be a fluid, he was right in proposing the nervous system delivered electricity to muscle tissue to make it move. We now know those neural impulses travel along specialized structures called axons that transport the electrical impulse generated by a neuron to other nearby neurons or muscle spindles that contract or expand the muscle. But Galvani's insight stood out, helping to establish modern neuroscience and demonstrating that the brain uses a known force to function. And if the brain uses electricity to operate and support human intelligence, can other devices using electricity achieve the same?

|  |  |  |
| --- | --- | --- |
|  | A close-up of a sword  Description automatically generated with low confidence |  |
|  | **Figure 3.2** - Galvani's drawings of electric stimulation to frog legs |  |

### 3.3 - The Loom (1801)

Invented in 1801, the Jacquard Loom uses a chain of punch cards to instruct a loom on how to make intricate textiles. Later, when designing his Analytical Engine, one of the first computers ever to be created (although never built), Charles Babbage was inspired to use the perforated cards to input commands and data into the engine. Mathematician Ada Lovelace went further, envisioning a general-purpose computing machine with applications beyond pure calculation. Later, Herman Hollerith also used the idea of punch cards to not only store information but also used it as input into a computing device, helping to move IBM into the computer age.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  | **Figure 3.3** - Diagram of the complex machinery of the Loom |  |

### 3.4 - Mary Shelley's Frankenstein (1818)

Galvani's work did not only advance science, but it also served as an inspiration to one of the most famous stories ever put on paper. Young Mary Shelley had learned of Galvani's findings, and she was intrigued about the moral implications that came with it. When reading her novel Frankenstein, or The Modern Prometheus, it is clear she saw a potential future where humans could create life. But Frankenstein is more than a tale of bringing the dead back to life. Its central theme is a question of ethics and morality. To bestow intelligence, sentience, and life to something comes with great responsibility. The creature brought to life by Frankenstein confronts his creator in dialogue to point this out: "*Yet you, my creator, detest and spurn me, thy creature, to whom thou art bound by ties only dissoluble by the annihilation of one of us. You purpose to kill me. How dare you sport thus with life*?" He adds a few moments later: "*Remember that I am thy creature; I ought to be thy Adam, but I am rather the fallen angel, whom thou drivest from joy for no misdeed*." As we advance our AI and understanding of genetics, we force ourselves to think critically about these topics.

|  |  |  |
| --- | --- | --- |
|  | Graphical user interface  Description automatically generated with medium confidence |  |
|  | **Figure 3.4** - Frankenstein, written by Mary Wollstonecraft Shelley (30 August 1797 – 1 February 1851) and first published in 1818, is considered one of the earliest examples of science fiction. Her novel is far more than a gothic horror story. It raises some critical questions and concerns about human responsibility when creating life, either through genetic engineering or artificial intelligence. |  |

### 3.5 - Modern Neuroscience (1860 – onward)

Modern neuroscience was born about a century later. The work of Golgi and Santiago Ramón y Cajal in staining and visualizing neurons led the way for the neuro doctrine: the scientific stance that all brain function depends on neurons and the connections between them. Lesion studies also supported this perspective: the specific loss of function that followed damage or specific brain areas such as Broca's aphasia.

|  |  |  |
| --- | --- | --- |
|  | A picture containing text, envelope  Description automatically generated |  |
|  | **Figure 3.5** - Cajal's beautiful drawings of neurons and their structure are true works of art, in addition to representing the massive leaps forward in our understanding of the nervous system. |  |

### 3.6 - Alan Turing (1936)

In 1936, Turing published his paper On Computable Numbers, introducing his idea of a Universal Computing Machine. In theory, such a hypothetical machine could be capable of any possible mathematical computation if an algorithm could be for it. His efforts are a reformulation of Kurt Gödel's paper on the limits of proof and computation from 1931. Of course, many of you know him as one of the scientists at Bletchley Park responsible for cracking the German encryption device known as the Enigma during the second world war. After the war, he continued to work in abstract mathematics and proposed what is now known as the Turing test, an attempt to devise some measure of real artificial intelligence. The Turing test is simple but sound. Human subjects interact with an AI through some terminal to hide any clues about the physical form of the AI. To Turing, the goal of AI is not necessarily to build something that looks and behaves like a human but can reason like a human when asked questions. The human does not know whether they are conversing with an AI or another human. Fooling the human into believing the AI is human means the AI has passed the Turing test. Turing himself did not see the massive impact his work had on humankind. He committed suicide after being prosecuted for homosexual acts by the same government he helped fight and defeat the Nazi forces. It wasn't until 2013 that Queen Elizabeth II posthumously pardoned him.

|  |  |  |
| --- | --- | --- |
|  | A person in a suit  Description automatically generated with medium confidence |  |
|  | **Figure 3.6** - Alan Mathison Turing (23 June 1912 – 7 June 1954) was an English mathematician, computer scientist, philosopher, and theoretical biologist. He is widely considered the father of theoretical computer science and artificial intelligence. Turing was highly influential in the development of theoretical computer science. Most importantly, he formalized the concepts of algorithms and computations with the Turing machine, considered a proto-model of a general-purpose computer. |  |

### 3.7 - Dartmouth Conference (1954)

The first use of the term Artificial Intelligence was at a conference at Dartmouth in 1954. It appears in the original proposal for the conference by John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon: "*We propose that a 2-month, 10-man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed based on 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 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.*"

|  |  |  |
| --- | --- | --- |
|  | Logo, company name  Description automatically generated |  |
|  | **Figure 3.7** - The first known usage of the word Artificial Intelligence was in 1954, at a conference in Dartmouth. |  |

### 3.8 - Eliza (1966)

One of the first AI applications was ELIZA, a program emulating a Rogerian psychotherapist. ELIZA is somewhat of a hoax. Rather than offering any profound insights into your life and problems, ELIZA just asks you to elaborate on things you said before. With our exposure to modern chatbots, we now know better. But when the original ELIZA first appeared in the 60s, many people mistook her for an actual human therapist.

|  |  |  |
| --- | --- | --- |
|  | A screenshot of a computer  Description automatically generated with low confidence |  |
|  | **Figure 3.8** - Dialogue between ELIZA and a subject. In its time, ELIZA could fool quite a few people into believing the AI was a human therapist asking questions through the computer terminal. It is unlikely that, in this current age, we would be fooled by it, given our exposure to chatbots and voice assistants. But then again, why are we currently fooled by what we believe are actual human beings interacting with us? |  |

### 3.9 - Diverging fields (1970)

In 1970 the AI field started to show a clear diverging. From one perspective, AI is a collection of well-established logic, rules, and conditions that could replace a human expert with the same domain knowledge. It is very much in line with the thinking about intelligence at the time, both in computing and cognitive psychology. On the other side, the idea is that intelligence is less well-defined and should mimic the known architecture that supports it: the brain. Artificial neural networks offered a new take on traditional AI thinking but hit a few critical roadblocks.

### 3.10 - AI becomes useful (1980)

In the 1980s, the expert system approach to AI became commercially viable. There were real-life situations in which this approach was initially highly successful. The models were very narrow in scope, which is where they work best. They do not handle ambiguity well. But for some big corporations, these models were beneficial even in this limited scope. Examples are banks deciding on extending credit lines or the automotive industry building systems for mechanics to diagnose engine problems.

|  |  |  |
| --- | --- | --- |
|  | A picture containing whiteboard  Description automatically generated |  |
|  | **Figure 3.9** - Knowledge-Based Systems, with the decision tree like if/then logic, were amongst the first successful applications of AI. It represents what has been dubbed 'Good Old-Fashioned AI,' where the knowledge and skills to perform a complex task (say, medical diagnosis) is extracted from human experts and formalized as rules implemented in the AI. Although this approach yielded some successful models, it falls short of the promise to mimic biological intelligence, and our AI should be able to learn, rather than be told, how to interact with the world. |  |

### 3.11 - A parallel timeline

In a parallel timeline, work on neural networks continued. McCulloch and Pitts are often credited with the first mathematical description and implementation of an artificial neural network, modeling these networks after electrical circuits. Around the same time, neuroscience started laying the foundations of how brains and neurons learn through experience. Most notable was Hebb's Organization of Behavior, in which he coined the phrase: neurons that fire together, wire together. In other words, neurons that are active simultaneously strengthen their connection. Further research focused on how patterns of specific connectivity can represent information, like those described by Rosenblatt, Hopfield, and Kohonen.

### 3.12 - Trouble in Paradise (the 1980s)

Research on neural networks started declining around the 80s for three reasons. First, the other arm of AI began to prove its worth and success. Funding, therefore, shifted away from neural network research. Second, there were practical limitations on building neural networks: the computers at the time were simply not designed to do any of that type of processing. They still are not technically speaking, but they are powerful enough to emulate such behavior. More importantly, a theoretical limitation significantly undermined neural networks' usefulness: the XOR separability problem. With a simple neural network of one input layer and one output layer, we can easily assign weights to the units to implement the basic logic gates used in conventional computing, such as an AND and OR gate. However, we cannot implement the XOR gate using the same neural network architecture because the XOR problem is not linearly separatable. For a neural work of two layers, the state space is 2D. The problem is that we simply cannot draw a line dividing the positive from the negative examples, something we could do for our OR and NOT problems.

|  |  |  |
| --- | --- | --- |
|  | Diagram  Description automatically generated with low confidence |  |
|  | **Figure 3.10** - The infamous XOR problem. The AND and OR logic gates can only need a single layer to transform the input into the desired output. However, the XOR logic gate is not linearly separable, meaning that you cannot draw a line through the state space with all the True outputs (black dots) on one side and all the False outputs (white dots) on the other side. |  |

### 3.13 - AI Winter (mid-80s)

In the late 80s, research entered the so-called AI winter. Neural networks fell short of their initial promise in not being able to perform all necessary logic operations without at least additional layers. But adding layers required a new learning algorithm, which still eluded researchers. Concurrently, knowledge-based systems proved to be too rigid. The systems proved hard to maintain, inflexible, expensive to build, and used hand-curated rules. In practice, human expert decision making more subtle and holistic than rules and simple probabilities.

### 3.14 - Revival of AI - 1980s

The AI field has gone through a series of revivals. First, computing power continues to grow and become cheaper and more accessible. At the same time, backpropagation became a critical innovation in training multilayer neural networks. It is slow but reliable and still used today, even though no biological counterpart exists in biological neural systems.

### 3.15 - Machine Learning as a statistical problem (1988)

After reformulating AI as a statistical problem, the diverging fields finally came together again. It became clear that machine learning and AI must adopt a probabilistic approach, where input signals provide evidence that, when combined, informs the output of our predictions. This idea came from work on machine translation, one of the early NLP (Natural Language Processing) algorithms investigated.

### 3.16 - Deep Blue (1997)

In 1997, Garry Kasparov was defeated in chess by Deep Blue, a chess-playing supercomputer at IBM. It brought AI into the public domain and was a big triumph. However, critics argued this wasn't actual artificial intelligence. Deep blue simply used brute force to compute the value over many sequences of moves. Deep blue never really did anything that is especially hard for a computer but comes naturally to a human. The Internet starts to take off at the end of the 90s and into the 2000s. People quickly realize it offers an ideal testing ground for algorithms and collaborations. But more importantly, it starts to produce a steady supply of easily accessible training data through its content. Shortly after, eCommerce and social media tech discovers the massive value of AI in understanding and predicting user behavior. Ever-increasing traffic and content generation allows for extraordinarily complex models. At the same time, machine learning needed to mature fast to handle data at the scale produced by Twitter and Facebook.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  | **Figure 3.11** - On the left, the moment Kasparov realized he had lost to IBM Deep Blue. On the right, Lee Sedol struggling against DeepMind in the game Go. During this challenge, DeepMind won 4 out of 5 games, Sedol won the remaining game. |  |

More recently, an AI playing a different board game made headlines worldwide. AlphaZero, developed by the AI research company DeepMind was a follow-up to their original AlphaGo. Both algorithms played human opponents in the game Go, with AlphaGo performing at a reasonable professional level. It was trained on games played by humans and thus developed some sense of human strategies and how to apply them. AlphaZero did things very differently. Instead of learning from human experts, it played itself millions of times. At first, most games were nonsense and did not even lead to a winner. But over time, AlphaZero developed its unique strategies and outperformed its predecessor. When AlphaZero finally played the highest-ranking Go players in the world, it not only won. It won using strategies never seen before in human players.

### 3.17 - A New Era of Computing (the early 2010s)

Computing resources and availability of data and algorithm continues to grow as cloud computing services begin to take a foothold. Google releases MapReduce, paving the way for large-scale computing. Although initially an innovation for data processing, its value to machine learning becomes shortly after. Especially useful for the continued research on neural networks, Graphics Processing Units (GPUs) turned out to be exceptionally well suited to do the heavy computational lifting during training,

### 3.18 - Deep Belief Networks (the 2010s – onward)

Snowballing ever more, computer hardware and online services become cheaper, faster, and more reliable. At the same time, new advances in learning paradigms, activation rules, and loss functions drastically speed up learning and pave the way for deep belief neural networks of ever-increasing complexity and depth. Big tech decides to go all-in on AI and provides the necessary resources to drive innovation forward. Academics now start working for Tech giants like Google and Facebook, moving away from traditional academic settings. Two scientists stand out as the founding fathers of this new era of neural networks, Geoff Hinton and Yann LeCunn. Their combined contributions have massively accelerated what neural networks can do.

|  |  |  |
| --- | --- | --- |
|  | A picture containing person, person, indoor, posing  Description automatically generated |  |
|  | **Figure 3.12** - Geoff Hinton and Yann LeCunn. Their combined work, and that of their students, lay the foundations of a significant push forward in neural networks. Neural networks now seem poised to replace conventional machine learning algorithms altogether, as they are versatile and can be designed to take on many different machine learning tasks. More complex neural networks are possible with advances in hardware, and these new networks considerably improve performance and generalizability. |  |

### 3.19 - Convolutional Neural Nets (the late 2010s)

After 2010, researchers introduced a new generation of neural networks specifically designed to do machine vision tasks. The Internet provides the massive amount of image data needed for training these extraordinarily complex models, and data set initiatives like ImageNet are introduced. The ImageNet challenge gave us AlexNet, the first of many more extensive neural networks for computer vision. Since AlexNet, many architectures have been published, with ever-increasing accuracies for classifying images and other complex input data.

|  |  |  |
| --- | --- | --- |
|  | Diagram  Description automatically generated |  |
|  | **Figure 3.13** - The AlexNet architecture. The AlexNet architecture has become somewhat of a standard in neural network design, primarily regarding processing images or video. The reason is that part of the architecture consists of convolutional layers. For reasons we will say later, such layers provide the ideal machinery for many image-based machine-learning tasks. |  |

### 3.20 - 2020s

We continuously see novel architectures, and the models are becoming increasingly complex. Generative models and general adversarial networks (GANs) have begun to take a foothold and show their capabilities and potential to the public through style transfer and deep fakes. New natural language processing models like GPT-3 allow for more realistic text summarization and production. In addition, the way we train AI and machine learning models has changed. Whereas current models still need large amounts of structured and labeled data, more recent models use reinforcement learning approaches. By trial and error and positive reinforcement, rather than relying on human expertise and well-defined errors, these models can learn and adopt completely novel strategies that sometimes are superior to even the best human experts. For example, AlphaGo Zero, a relatively simple machine learning algorithm, trained itself by playing itself in millions of games of Go, each time reinforcing specific strategies that led to wins, even if those wins during the early phases of training came at random. Reinforcement learning offers a more holistic approach to solving a real-world problem. AI now learns more fundamental and conceptual aspects of the problem rather than evaluating single moves or actions.

### 3.21 – The Bottom Line, the singularity, and beyond

With the exponential growth in AI capabilities, the question asked by many science fiction writers suddenly became quite relevant. At what time will artificial intelligence match and overtake our intelligence? Are our algorithms subject to the same principles, whether ethical, moral, or legal? Should we be concerned with the possibility of rendering ourselves obsolete? Unfortunately, years of science fiction writing has us perhaps fearing and focusing on the wrong aspect of AI. Too often do we equate the dangers of AI with robots going rogue, becoming sentient, and disobeying human orders. Although not impossible, this will unlikely happen soon. Instead, we have ignored the disembodied AI, processing massive amounts of data in the background, making decisions that are difficult to justify even by those who build and train the algorithms, and mining the data points of our lives, turning our livelihood into a commodity. Truly dangerous AI doesn't come in cyborg form. It presents itself almost like a corporation, where all that truly matters is the bottom line. It is encouraging that we, the users, are actively questioning this approach, and demanding insights into how our data is collected, stored, and used to turn a profit. At the same time, a whole new industry of ethical AI has come into existence to regulate and oversee the fairness of the algorithms we use. But how do we measure fairness? Correctness? Bias? We will return to this at the very end when hopefully, your insights into the inner workings of machine learning will give you a better context for answering that question yourself. We won't bias you here with our stance before you have that chance to do so.

### 3.212- References

* Wooldridge, M. (2021). A Brief History of Artificial Intelligence: What It Is, Where We Are, and Where We Are Going. United States: Flatiron Books.
* Cruse, M. (2021). A History of Science: From Agriculture to Artificial Intelligence. United Kingdom: Arcturus Publishing.
* Turing, S. (2012). Alan M. Turing: Centenary Edition. United Kingdom: Cambridge University Press.
* Ferguson, K., Pumperla, M. (2019). Deep Learning and the Game of Go. United States: Manning.
* Norvig, P., Russell, S. J. (2016). Artificial Intelligence: A Modern Approach. United Kingdom: Pearson.
* Kasparov, G. (2017). Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins. United Kingdom: Public Affairs.
* Zhou, Y. (2017). Alphago Vs Lee Sedol: The Match That Changed the World of Go. CreateSpace Independent Publishing Platform.
* Shelley, M. W. (1992). Frankenstein (Wordsworth Classics) (Wordsworth Classics). United Kingdom: Wordsworth Classics.
* Koester, J. D., Kandel, E. R., Mack, S. H., Siegelbaum, S. A. (2021). Principles of Neural Science, Sixth Edition. United States: McGraw-Hill Education.
* Minsky, M., Papert, S. A. (2017). Perceptrons: An Introduction to Computational Geometry. United States: MIT Press.
* Alan Turing: His Work and Impact. (2013). Netherlands: Elsevier Science.
* Newman, E., Dubinsky, J. M., Araque, A., Swanson, L. W. (2017). The Beautiful Brain: The Drawings of Santiago Ramon Y Cajal. United States: ABRAMS.