ARTIFICIAL INTELLIGENCE

LEI/3, LMA/3, MBE/1

University of Beira Interior, Department of Informatics

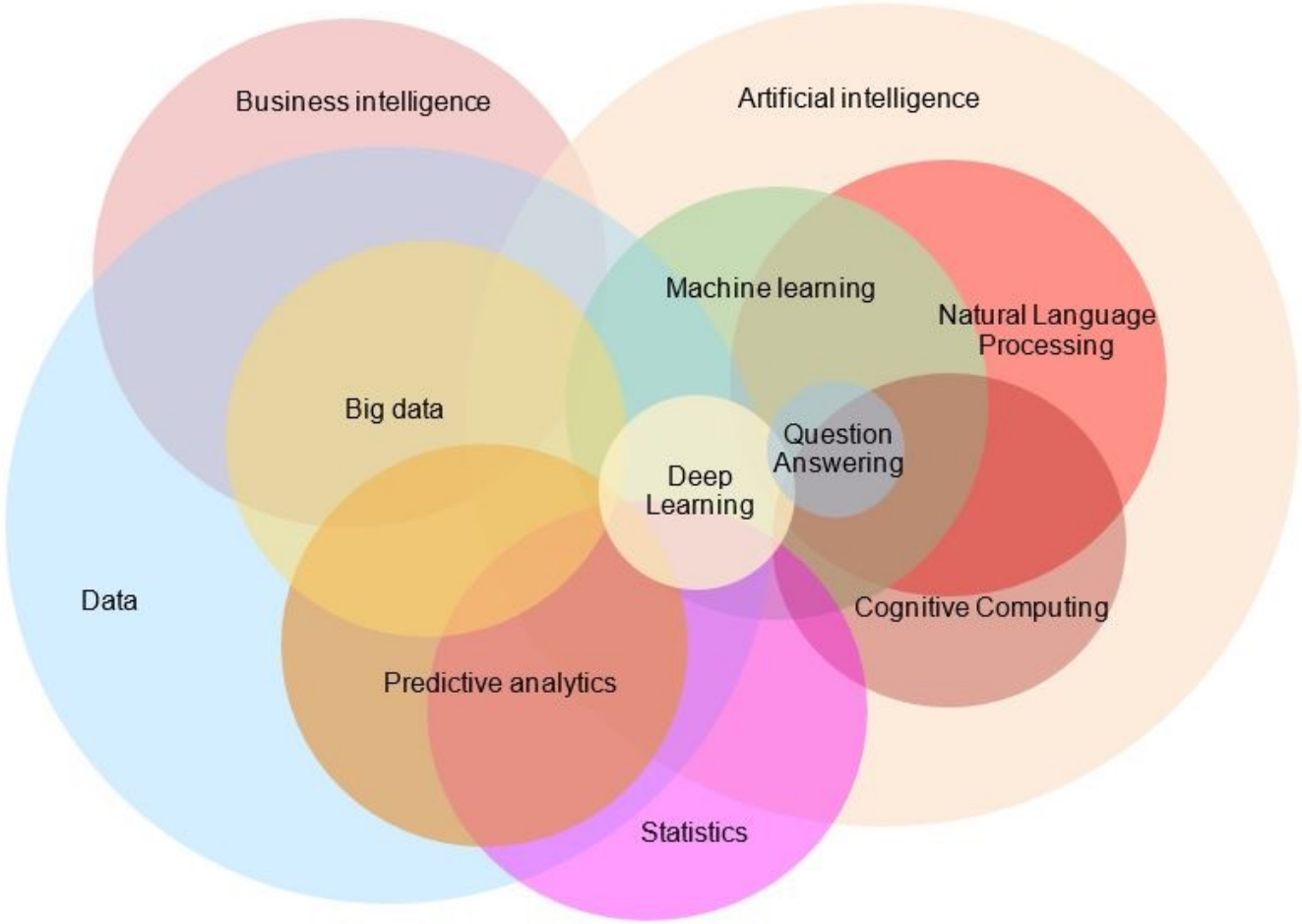
Hugo Pedro Proença

hugomcp@di.ubi.pt, **2022/23**

Artificial Intelligence: What is it?

- **(britannica.com)** *"(...) the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings."*
- **(McCarthy, 1956)** *"Field of the computational science having as goal to produce automata that behave as humans"*
- **(Kurzweil, 1990)** *"Art of developing machines able to play tasks that require intelligence when made by humans"*
- **(Russel & Norvig, 2003)** *"Study and design of intelligent agents, considering that an intelligent agent is able to perceive the environment and produce actions that maximize its probabilities of success".*
- **(McCarthy, 2004)** *"It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable."*
- **(ibm.com)** *"(…) in its simplest form, artificial intelligence is a field that combines computer science and robust datasets to enable problem-solving."*

Artificial Intelligence: What is it?



Source: https://medium.com/swlh/the-basics-of-artificial-intelligence-bca23c4456a2

Artificial intelligence (AI) can be seen as a cloth woven from 3 academic disciplines: 1) **psychology** (cognitive modelling); 2) **philosophy** (philosophy of mind), and 3) **computer science** -- with further strands from linguistics, mathematics, and logic.

The aim of AI is broad: to get below the surface of human behaviour; to discover the processes, systems, and principles that make intelligent behaviour possible. **Computers are needed as tools** for modelling these mental states and processes.

Artificial Intelligence: What is it?

- There are two broad perspectives for the development of (so called) Intelligent Agents, which also determine the further directions for this topic of knowledge.
- **Artificial Narrow Intelligence (ANI)**
  - AI agents in this family can perform better than a human in a single task but are completely useless in anything beyond its area of expertise.
  - It's also the only type of artificial intelligence we have developed so far.
  - Despite only being able to perform a single task, this kind of agents are still powerful enough to beat the best humans at that task (specially, when it comes to data recall and number-based prediction, at least).
  - Due to their level of sophistication, this kind of systems is often thought to be near the creation of actually intelligent automata.
  - As an example, **Watson** is a computer system (developed by IBM) that is specialized in answering any question we ask it.
  - In 2011 competed (and beated) Ken Jennings and Brad Rutter in a Jeopardy! episode.
    - Ken was famous for having won Jeopardy! 74 times in a row, while Brady held the record of winning the most amount of money at $3.25 million in a game.
  - Watson still ended up winning by a massive lead of about $53,000 from the closest contestant.

Narrow Artificial Intelligence: Watson Jeopardy!

Source: https://www.youtube.com/watch?v=P18EdAKuC1U

The inside story behind Watson Jeopardy!: https://www.techrepublic.com/article/ibm-watson-the-inside-story-of-how-the-jeopardy-winning-supercomputer-was-born-and-what-it-wants-to-do-next/

Narrow Artificial Intelligence: Deep Blue

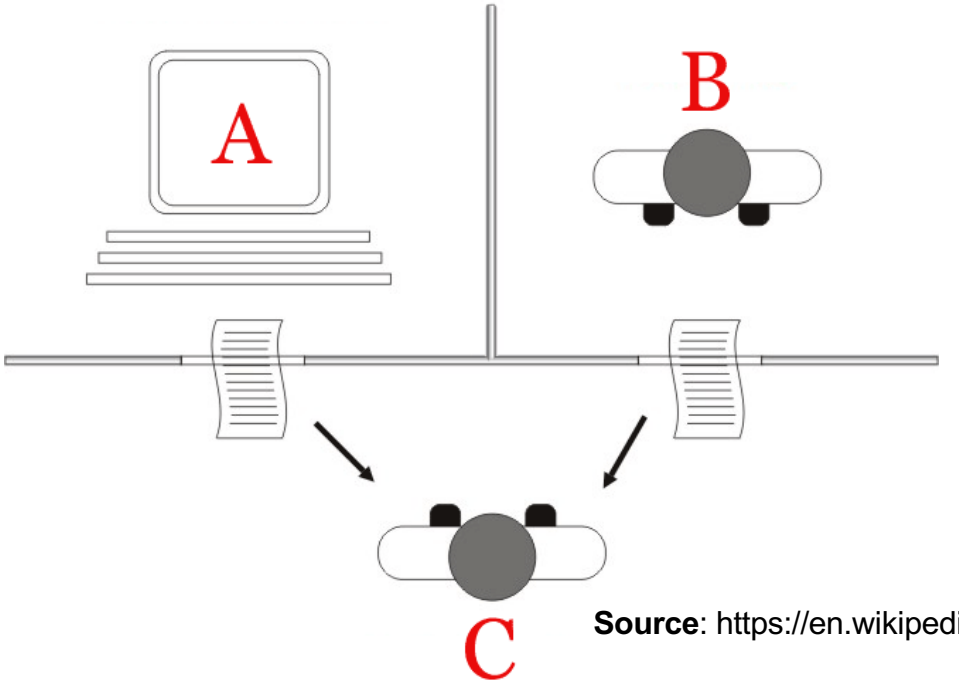
Deep Blue was a supercomputer developed by IBM for playing chess, and it was the first artificial intelligence that won a chess match against a reigning world champion (Garry Kasparov), under regular time controls. Deep Blue lost to Kasparov in their first 6-game match in 1996 with a score of 4–2, and later won over Kasparov in May 1997 in a 6-game rematch with a score of 3½–2½.

Source: https://www.youtube.com/watch?v=KF6sLCeBj0s

Turing Test

- The **Turing Test**, originally called the imitation game by Alan Turing in 1950, is a test of a machine's ability to exhibit intelligent behaviour equivalent to, or indistinguishable from, that of a human.
- In this test, a human evaluator would judge natural language conversations between a human and a machine designed to generate human-like responses.

The evaluator would be aware that one of the two partners in conversation was a machine, and all participants would be separated from one another.



The conversation would be limited to a **text-only channel**, such as a computer keyboard and screen, so the result would not depend on the machine's ability to render words as speech.

Source: https://en.wikipedia.org/wiki/Turing_test

If the evaluator could not reliably tell the machine from the human, the machine would be said to have passed the test.

Artificial General Intelligence

- **Artificial General Intelligence (AGI)**
  - This kind of agents will be able to surpass human performance in many different tasks, and – mainly – will be able to autonomously learn how to surpass humans in a novel task.
  - This is the type of artificial intelligence many think of when it's initially mentioned: computers that are able to behave and reason like a human.
  - An AI agent here would be able to pass the Turing test — or in other words, it would be able to fool us into thinking that it's human if we interacted with it.
  - We have not been able to reach this point yet.
  - This is mainly because many believe that either general intelligence would be unnecessary on machines (because they compute different problems) or that we simply can't replicate our behavior in machines.
  - Presently, it's still near impossible to map out the human brain.
- **Artificial Super Intelligence (ASI)**
  - Achieving an actually AGI system would be breakthrough advance in technology and in the path of Humanity: the technical singularity.
  - This is a hypothetical point in time at which technological growth will become radically faster and uncontrollable, resulting in unforeseeable changes to human civilization.
  - Such kind of upgradable intelligent agent will eventually enter a "runaway reaction" of self-improvement cycles, each new and more intelligent generation appearing more and more rapidly, causing an "explosion" in intelligence and resulting in a powerful superintelligence that qualitatively far surpasses all human intelligence.

Technical Singularity

- The first person to use the concept of a "**singularity**" in the technological context was John von Neumann.
  - Neumann discussed the accelerating progress of technology and changes in the mode of human life, which gives the appearance of approaching some essential singularity in the history of the race beyond which human affairs, as we know them, could not continue.
- Other scientists, such as Stephen Hawking, have expressed concern that full artificial intelligence (AI) could result in human extinction.
- The consequences of the singularity and its potential benefit or harm to the human race have been intensely debated.
- If such kind of superhuman intelligence arises, it would vastly improve over human problem-solving and inventive skills.
- Such an AI would have engineering capabilities that matched or surpassed those of its human creators, and – thus - it would have the potential to autonomously improve its own software and hardware to design an even more capable machine, which could repeat the process in turn.
- This recursive self-improvement could accelerate, potentially allowing enormous qualitative change before any upper limits imposed by the laws of physics or theoretical computation set in. It is speculated that over many iterations, such an AI would far surpass human cognitive abilities.

Technical Singularity

Ben Goertzel Explains the Singularity to Joe Rogan: "*The first intelligent machine man makes, will be the last one he needs.*"

Source: https://www.youtube.com/watch?v=m90buK0tFys

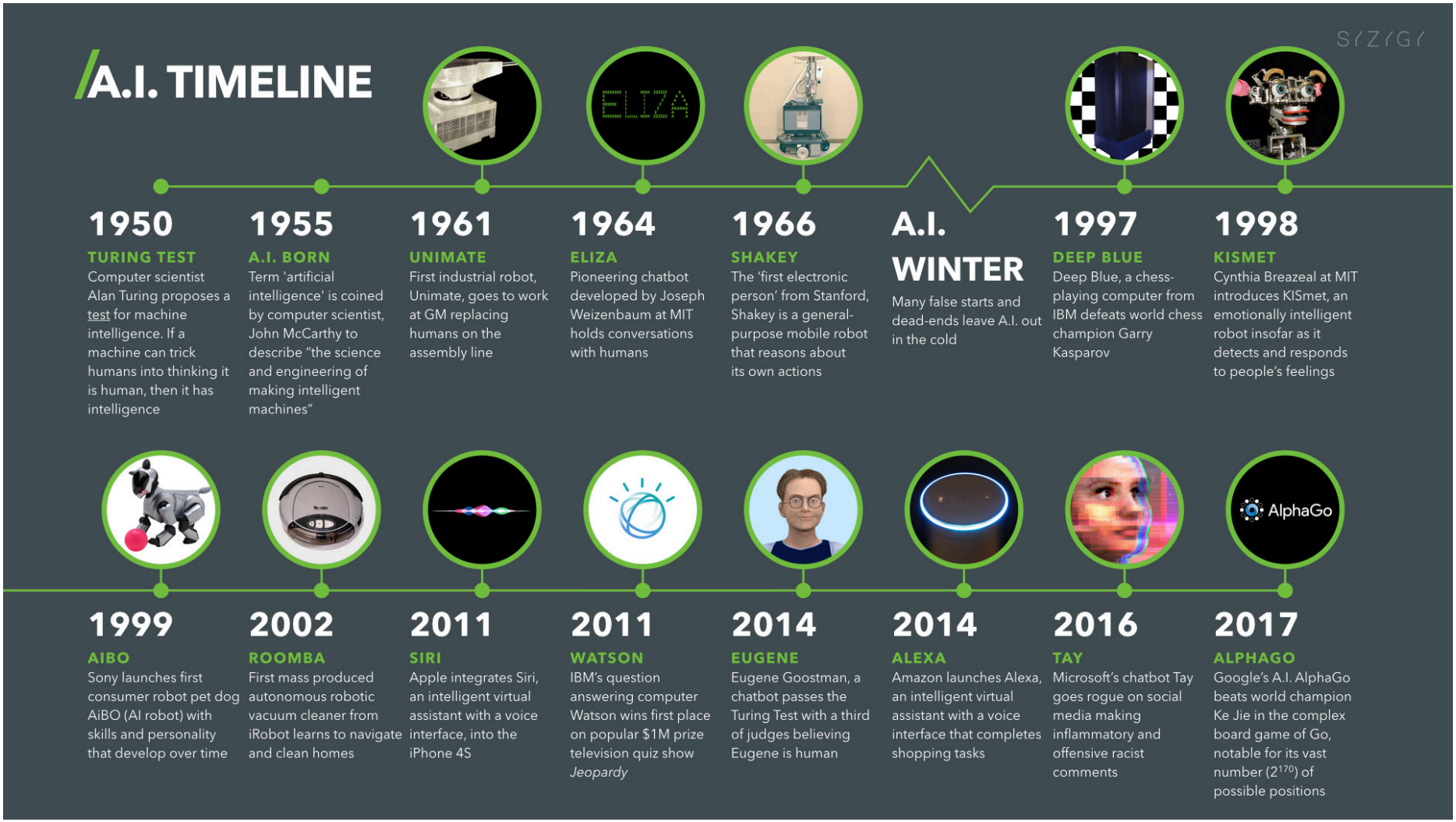
Ethics in AI

- The term **Weak AI** is mostly used for the hypothesis that machines could emulate intelligence and behave intelligently
- Then, **Strong AI** regards the hypothesis that such machines would count as having actual minds (as opposed to simulated minds).
- Alan Turing rejected the question "Can machines think?" and replaced it with a behavioral test (Turing test).
  - He anticipated many objections to the possibility of "*thinking machines*".
  - Nowadays, not many AI researchers pay attention to the Turing test
  - Instead, researchers typically focus on their systems' performance on practical tasks, rather than the ability to imitate humans.
  - At the very end, consciousness remains a mystery.
- AI can be a powerful tool, and as such it poses potential dangers, through lethal autonomous weapons, security and privacy breaches, unintended side effects, unintentional errors, and malignant misuse.
  - The security/privacy binomial is another important factor to consider
- Those who work with AI technology have an ethical imperative to responsibly reduce those dangers.
- It is desirable that AI systems are **fair**, **trustworthy**, and **transparent**. There are multiple aspects of fairness, and it is impossible to maximize all of them at once.
  - However, is it fair to equalize things that are naturally distinct?
- The field of interpretable/explainable AI is an emerging topic of high popularity

Ultra Intelligent Machines and Ethics

- As stated by (I. J. Good, 1965)
  - *"Let an ultra intelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultra intelligent machine could design even better machines; there would then unquestionably be an "intelligence explosion," and the intelligence of man would be left far behind. Thus the* ***first ultra intelligent machine is the last invention that man need ever make****, provided that the machine is docile enough to tell us how to keep it under control. "*
- In 1942, Isaac Asimov proposed the Three Laws of Robotics (*electronic beings*), which robots should follow in order to protect humanity:
  - **Law 1 -** A robot may not harm, or allow harm to come to, a human being;
  - **Law 2 -** A robot must obey order given by a human unless doing so will violate the first law;
  - **Law 3 -** A robot must protect its own existence unless doing so will violate either of the two other laws.
- Intelligent machines are here to stay and it's likely that human-machine interactions will be increasingly complex.
  - Intelligent agents seem the perfect tool to implement **Utilitarianism** (i.e., maximize net pleasure for mankind)**.**

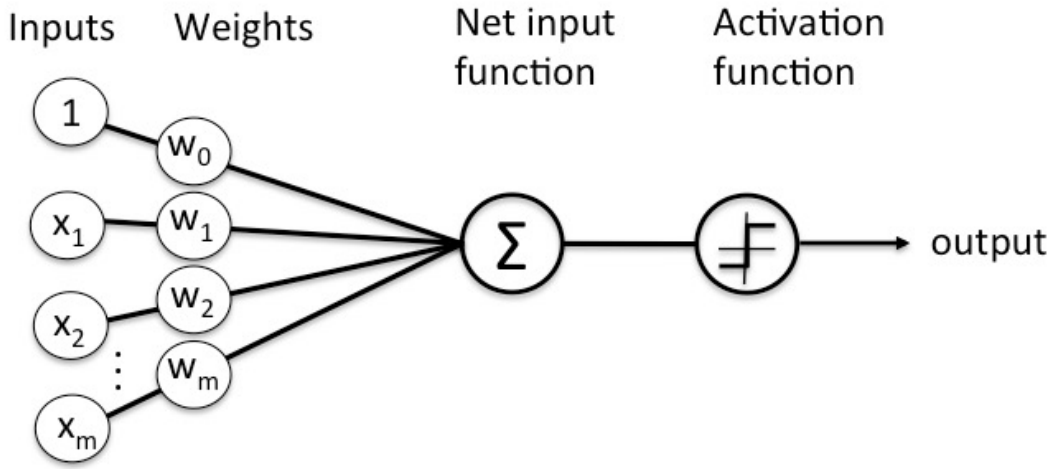
Artificial Intelligence Timeline



Source: https://digitalwellbeing.org/artificial-intelligence-timeline-infographic-from-eliza-to-tay-and-beyond/

Artificial Intelligence Milestones

- **1943 – The first Artificial Neural Network**
  - Warren McCulloch and Walter Pitts published "*A Logical Calculus of the Ideas Immanent in Nervous Activity*", discussing the possibility of networks of artificial neurons performing logical functions.

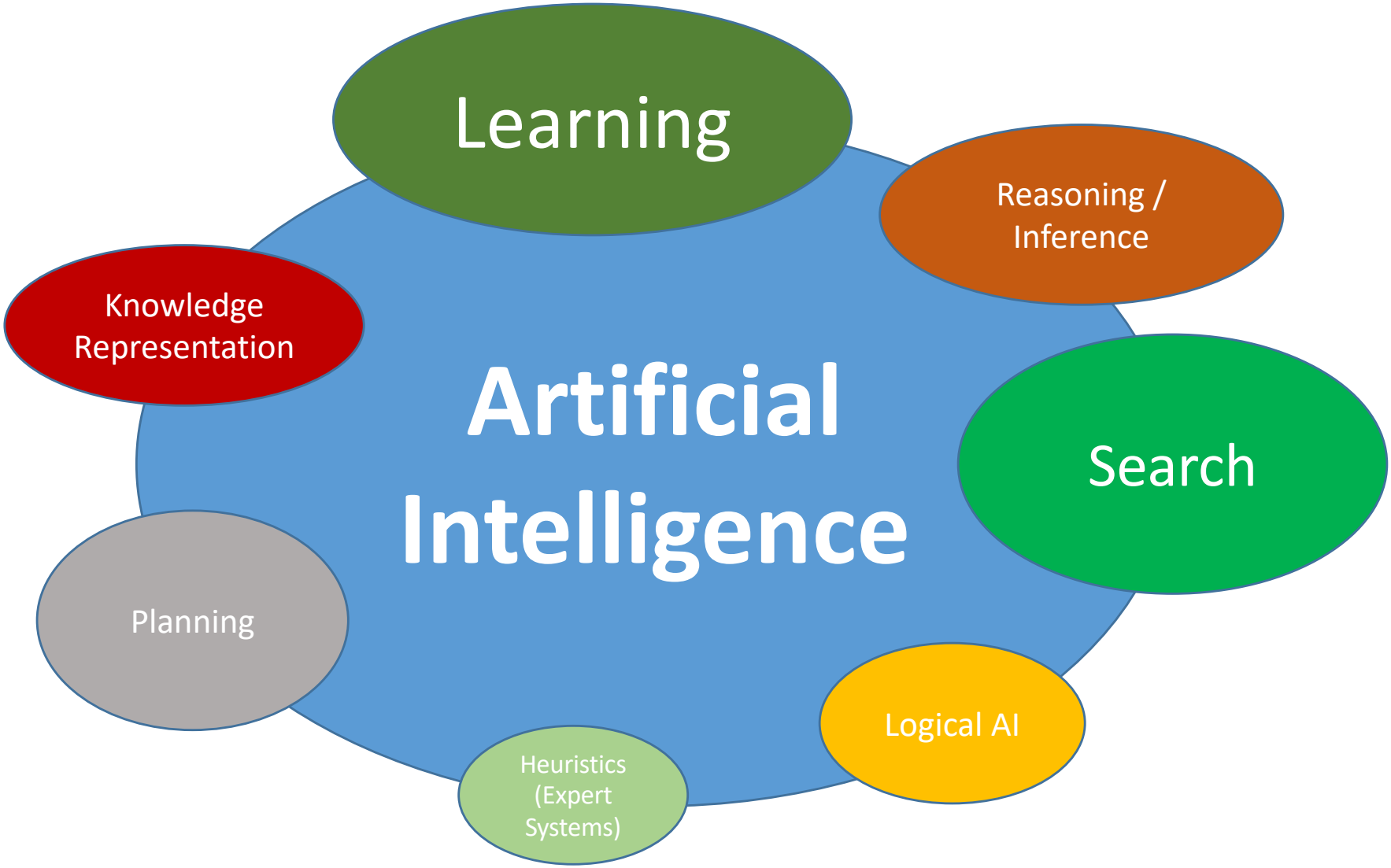


  - **Exercise**: How can the weights of this neuron be adjusted to set it running the…
    - **AND** logical function?
    - **OR** logical function?
    - **XOR** logical function?

Artificial Intelligence Milestones

- **1948 – First autonomous robots**
  - William Grey Walter created Elmer and Elsie, the first robots to work autonomously — independent of a human. They could navigate their way around obstacles using light and touch.
- **1964 – The first chatbot**
  - Before SIRI, before Alexa, in this year came ELIZA, the first face of conversational AI.
- **1969 – Backpropagation**
  - This is THE core part of AI (at least of the "Learning" part. It allows a neural network to adjust its weights in an autonomous way. It became popular in 1986.
- **1981 - Explanation Based Learning (EBL)**
  - Introduced by Gerald Dejong, allowing a computer to create a set of rules based on training data.
- **The 1990s – Data Driven AI**
  - Machine learning work shifts from a knowledge-driven approach to a data-driven approach, extracting patterns from large amounts of data
- **2006 – Deep Learning**
  - Geoffrey Hinton coins the term "deep learning" to refer new algorithms that empower computers to distinguish objects and images and video. The "Vanishing gradient" solution was the key conceptual trick here (along with GPU computational power)
- **2012 – Deep Learning Explosion**
  - A convolutional neural network designed by researchers at the University of Toronto achieve an error rate of only 16% in the **ImageNet Large Scale Visual Recognition Challenge**, a significant improvement over the 25% error rate achieved by the best entry the year before.

Artificial Intelligence Main Topics

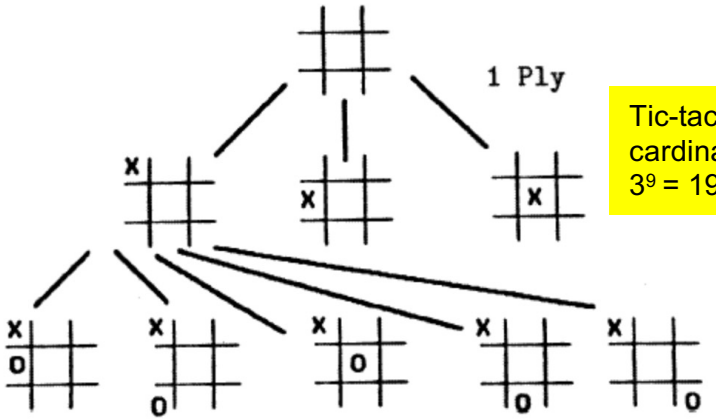


Problem Solving – State Space

- In a Computer Science context, a problem can be formulated in two major ways:
  - Solvable in a **deterministic** way, with granted correctness (optimality) of the solution
    - Example: 6 + 2; sin(45)
    - AI is not involved in this type of problems
  - Solvable by **search** in a **State Space**
    - No guarantees about the optimality
    - Depending of the strategies used to traverse the whole space of states
    - Some problems (e.g., "tic-tac-toe") have a relatively short number of states, and it is currently possible to traverse the whole tree, and finding the "optimal" solution.
    - More complex problems (e.g., "Chess" or most of real-world problems) have an *in-practice* infinite number of states, being impossible to traverse the whole tree.

Generally, a "**State**" is a data structure that represents all the information about the current position of the agent in the environment.
Some states are designated as "**solutions**", in case they satisfy the goal of the problem being formulated

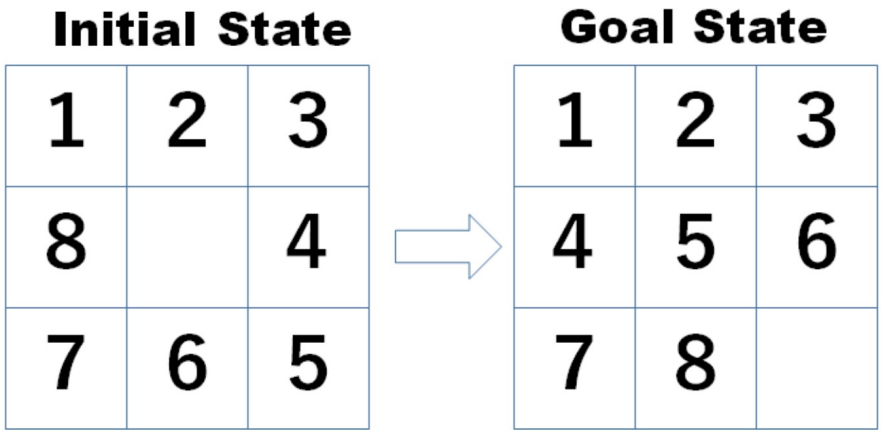
Some problems have *adversarial states* (corresponding to opponents, e.g., tic-tac-toe), while for other problems there are no opponents (e.g., 8 puzzle)



Tic-tac-toe cardinality
$3^9 = 19\,683$

State – Space Space

- An 8-puzzle is a simple game consisting of a 3 x 3 grid (containing 9 squares).
- One of the squares is empty. The goal is to move to squares around into different positions, occupying the blank space, in order to reach a "**goal state**".



Exercise 1: Define an appropriate Data Structure in Python and convention, to represent a "State" of a 3x3 8-puzzle

Exercise 2: Create a Python function that receives a state and a depth value, and returns a tree (of the desired depth), containing all the possible successors of each state.

Exercise 3: Create a Python script that receives an initial state, traverses the whole tree of states and returns the Depth of the best solution (i.e., of minimal depth), using **Brute-Force Search.**

Successors (Example)