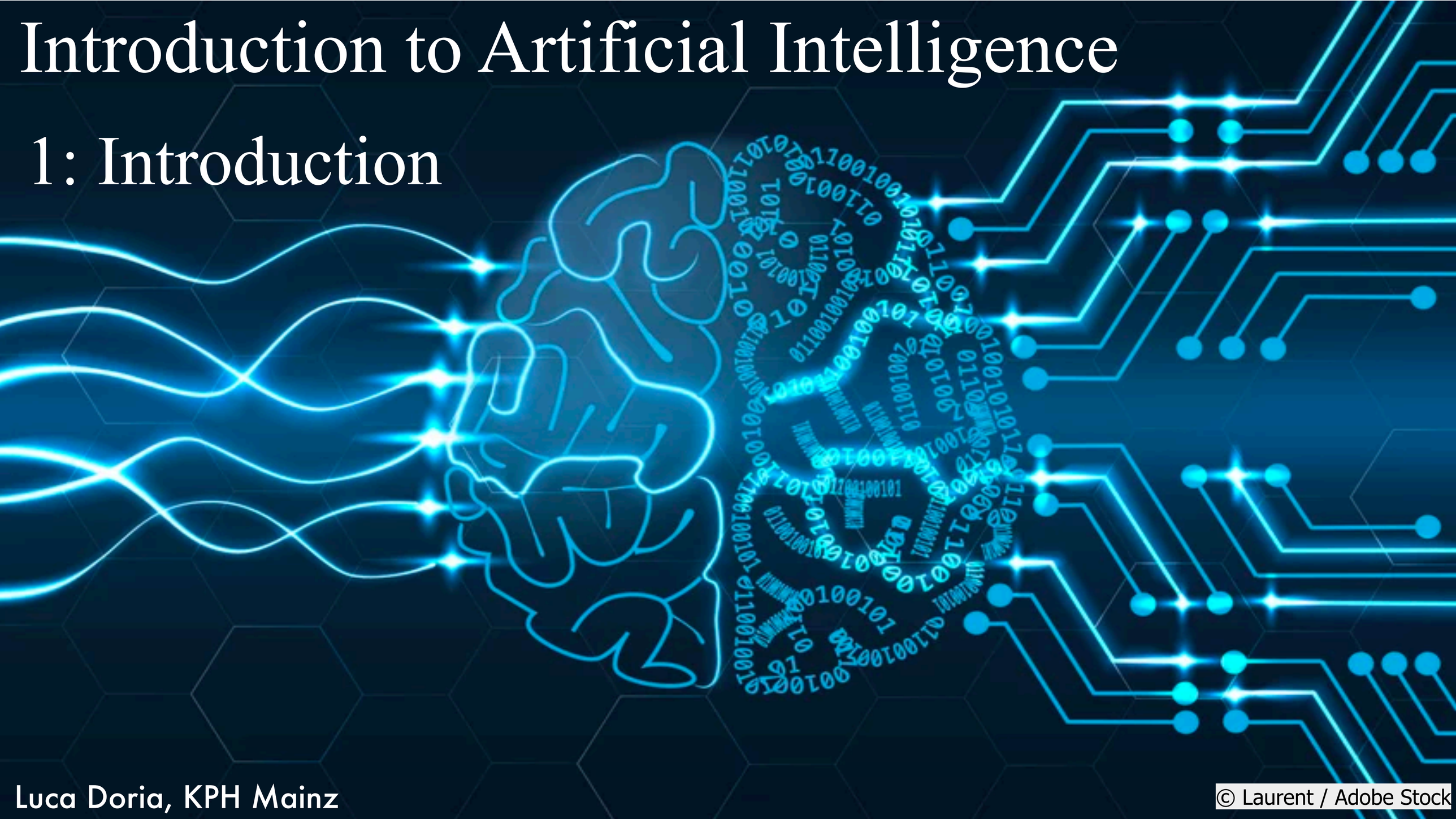
Introduction to Artificial Intelligence

1: Introduction



Luca Doria, KPH Mainz

Artificial Intelligence: A definition (?)

Historically, many different definitions were considered with emphasis on different aspects.
A simple classification scheme is:

	Humanly	Rationally
Thinking	**Knowledge of human "thinking" needed** How to understand it ? - Introspection - Psychology experiments - Brain Imaging —> Cognitive sciences	**Implement:** - Logic reasoning - Probabilistic reasoning
Acting	**(Total*) "Turing Test"-winning machine** Needs: - Natural language processing - Knowledge representation (storage) - Automated reasoning - Machine learning (adaptation) - Computer vision and robotics*	**Agent-based** - Machines that "do the right thing" - Mathematically defined rationality - Can interact with environment - Can store information - Can cope with imperfect knowledge - …

Artificial Intelligence: Ask an AI!

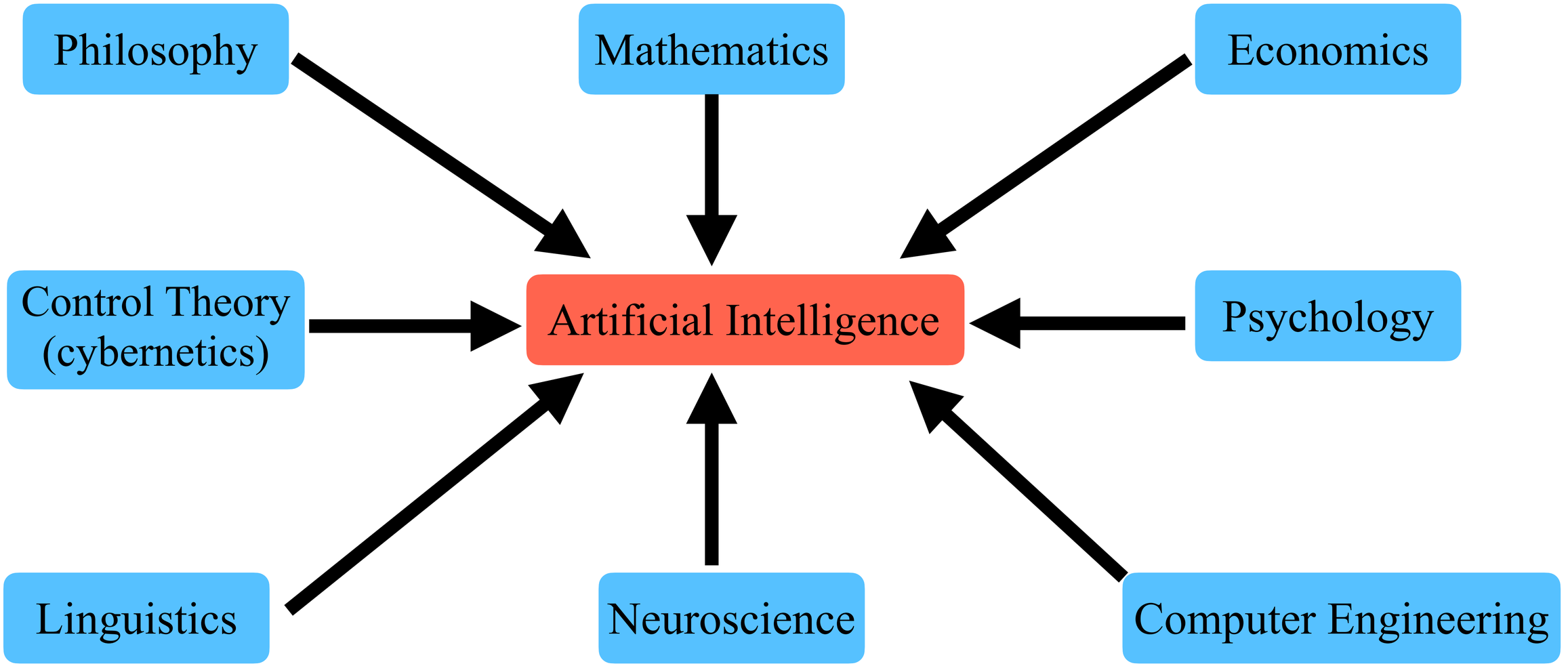
You

What is artificial intelligence?

ChatGPT

Artificial Intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think and mimic human-like cognitive functions. These machines are designed to perform tasks that typically require human intelligence, such as learning, reasoning, problem-solving, perception, speech recognition, and language understanding.

Contributions to AI from other fields



Philosophy

Likely the first field posing fundamental questions for AI:

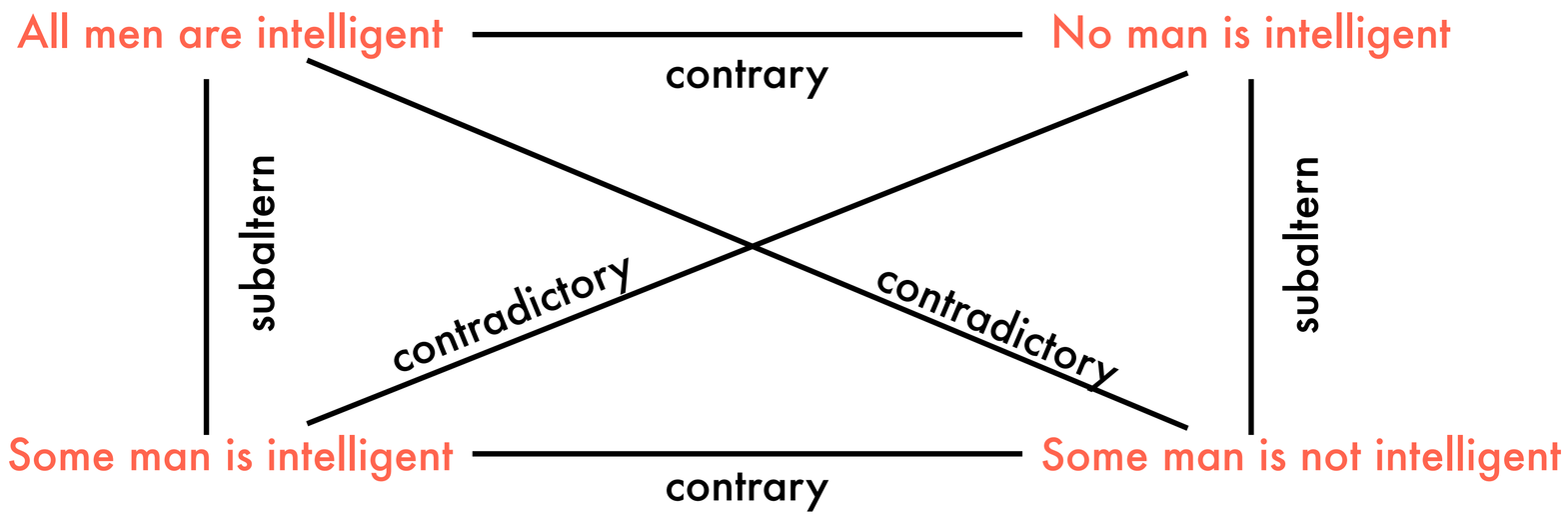
- What is the relationship between mind and physical brain?
- How do you define “mind”?
- What are the rules (if any) of (rational) thought?
- What is knowledge?
    - Where is it?
    - Where does it come from?
    - How is it organised (knowledge representation)?

Philosophy

Aristotle (384-322 BC)

Focus on the rational thought.
The founder of logic: first to develop a system for deriving conclusions from (true) premises (the “syllogisms”).

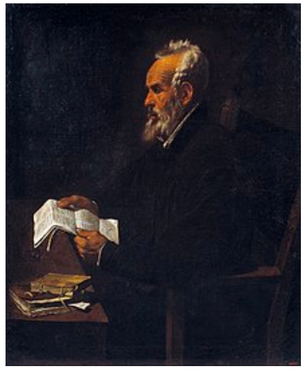
The famous “square of oppositions”

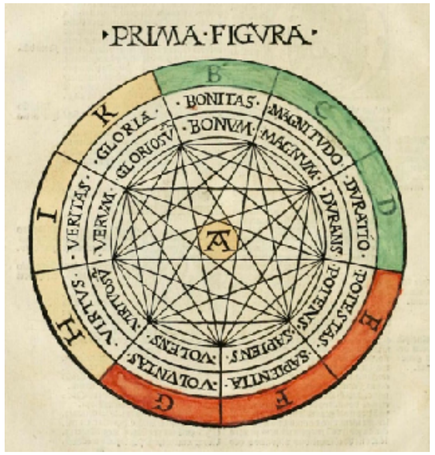


Philosophy

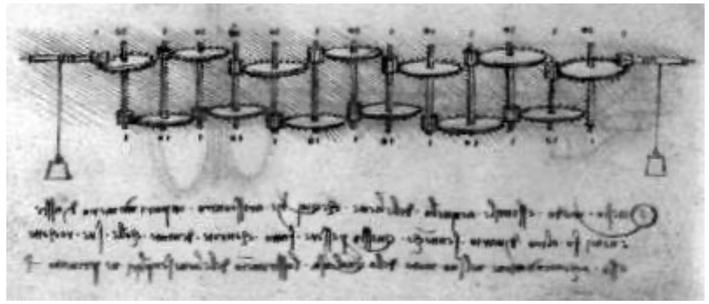
Many attempts to automatise the rules of rational thought.
Towards the idea of artificially reproduce the rational part of the mind.

Ramon Llull (1232-1315)





Leonardo da Vinci (1452-1519)



Leonardo's mechanical calculator, realised practically only later.

Wilhelm Schickard (1592-1635)

First ever constructed mechanical calculator

Blaise Pascal (1623-1662)

The "Pascaline"

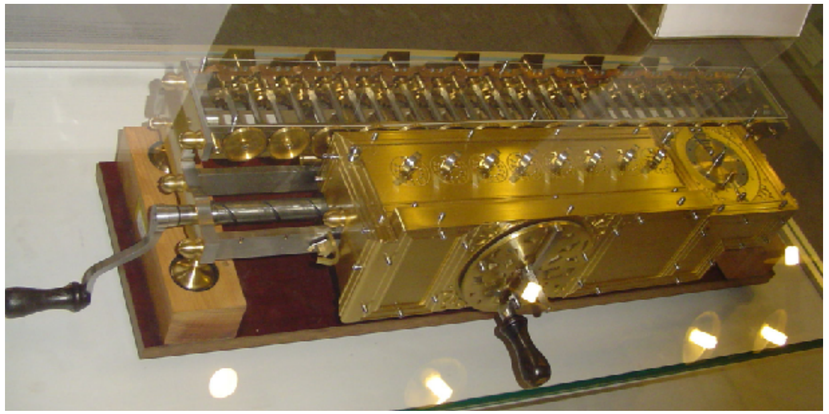
Philosophy

Rene' Descartes (1596-1650)

If the mind works like a machine there is not space for free will (materialism).
Proposed dualism: at least part of the mind is not physical, escaping mechanical laws.

Gottfried Wilhelm Leibniz (1646-1716)

Built a mechanical device which could work on concepts instead of numbers.
Convinced that a rational argument should convince any rational mind, he posed that in the future, once we will know the "rules" of though, every dispute will be settled not just
saying "let's discuss", but instead "let's calculate" (calculemus!).



Leibniz's Arithmetic machine, which can perform for the fist time +,-,x,/ operations.

(Reprodution from the Deutsches Museum).

Mathematics (and Th. Computer Science)

The logic of Aristotle slowly evolves to mathematical logic:

- Traces of the process in Greece, India, China
- George Boole (1815-1864): formal first order logic
- Gottlob Frege (1848-1925): first order logic
- Kurt Gödel (1906-1978): Incompleteness Theorem
- Alan Turing (1912-1954): Computability
- Stephen Cook (1939-), Leonid Levin (1948-), Richard Karp (1935-): Complexity Theory

Sometimes we do not know everything we need for a formal rational conclusion:
probability plays key role:

- Girolamo Cardano (1501-1576)
- Blaise Pascal, Pierre Fermat (1601-1665), J. Bernoulli (1654-1705), Laplace,…

Mathematics (and Th. Computer Science)

Propositional and Predicative logic are complete…
but surprisingly:

Gödel's First Incompleteness Theorem:
Every consistent (contradictions cannot be derived) formal system able to describe the natural numbers will contain unprovable truths (undecidability).

Gödel's Second Incompleteness Theorem:
Every consistent formal system able to describe the natural numbers cannot prove its consistency.

Another formulation (Turing): there are problems which cannot be solved algorithmically.
Or mathematically: non-computable functions exist!

An uncomputable function

The "Busy Beaver" Function:

BB(n) = the maximum number of steps a n-states Turing machine can make starting form an initially blank tape before halting.
BB(n) actually grows FASTER than ANY function you can write!

BB(1) = 1
BB(2) = 6
BB(3) = 21
BB(4) = 107
BB(5) > 47×10^6
BB(6) = ? (*)

Why?
Suppose exists f such that f(n)>BB(n) for every n.
Therefore, we can consider a Turing machine M with n states and simulate it up to f(n) steps.
If M didn't halt before, it will never do, since f(n) is larger than the maximum steps possible. But this would give us a way to understand if M halts or not, contradicting Turing's (and Gödel's) theorem.
Therefore, f cannot exist.

(*): the exact numbers depend of the exact TM definition

Consequences for AI?

If AI is based on a fixed formal system, Gödel's theorems imply limitations.

Should a computer work in a fixed formal system?

Is human intelligence algorithmic anyway?

Is there "something" that distinguishes the brain (the mind?) from a computer?
If there is, we cannot simulate our mind on a computer (on a Turing machine).

Are the limits related to the impossibility to compute or to computational complexity?

Are there some limits from physics (implications from the Church-Turing thesis)?

Economics

The economic activity is carried out by agents supposed to have some kind of "intelligence".
In the few-hundreds of years of its existence, the economic theory modelled it in different ways.

One example is utility maximisation. There exists a function called utility which all the agents present in an economic environment try to optimise.

The task is in general quite complex:

- You may want to maximise your money, but also your free time
- Other agents might be in competition with you
- Agents operate with finite resources
- Rules (like laws) can change due to government decisions, wars, other unexpected events.
- …..

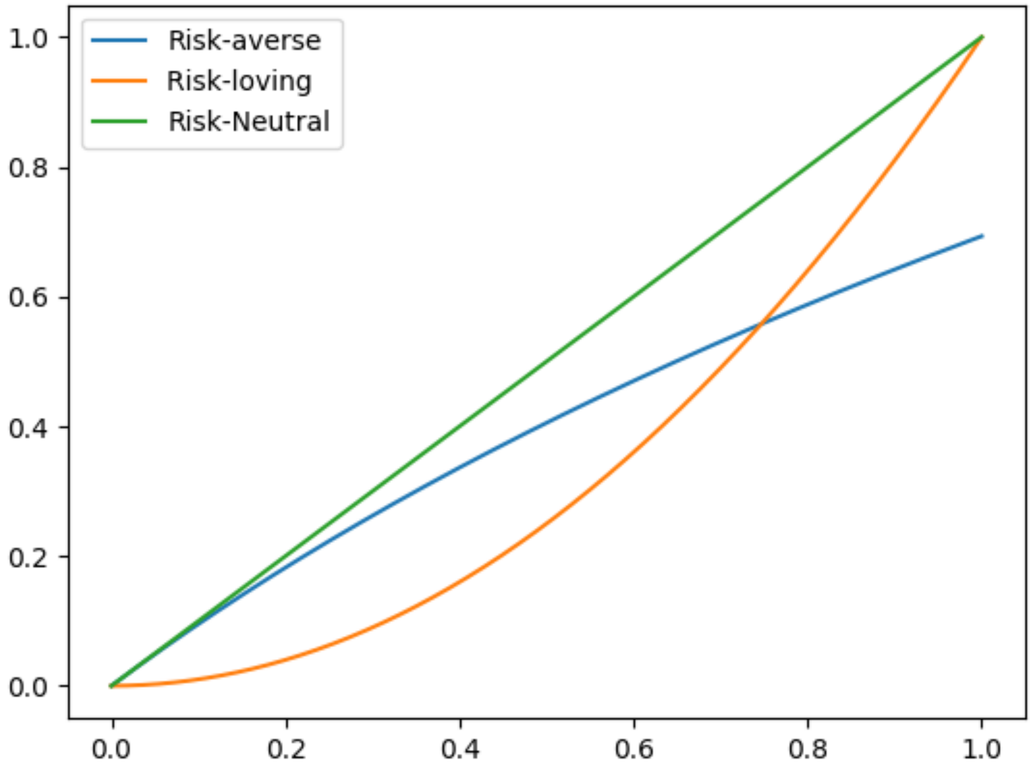
Economics

The **utility** is an abstract value that an economic agent gets from a preference:

Discounted utility: $U_t = \sum_{t=0}^{\infty} \beta^t u(c_t)$

β : discount factor

- Risk-averse, with a concave utility function;
- Risk-neutral, with a linear utility function, or;
- Risk-loving, with a convex utility function.



Economics

Example: Arrow-Pratt relative risk aversion utility:

Concavity/convexity : proportional to second derivative
Independence from affine transformations: divide by first derivative

$$\boxed{RRA = -c\frac{u''(c)}{u'(c)}}$$

If we consider a constant risk aversion:

$$RRA = k = -c\frac{u''(c)}{u'(c)} \quad \Rightarrow \quad u'' + \frac{k}{c}u' = 0 \quad \Rightarrow \quad \boxed{u(c) = \frac{c(t)^{\theta-1}}{1-\theta}} \qquad k = \frac{1}{\theta}$$

Neuroscience

Study of the nervous system, focus on the brain.

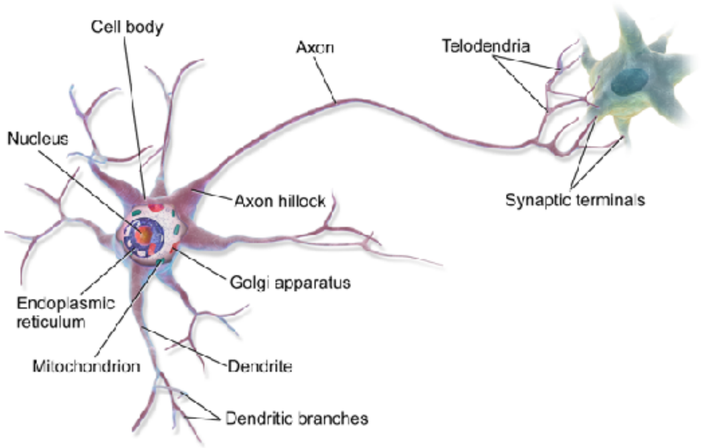
Only in the 18th century, the brain was recognised as the place where thought and consciousness are.

Some fundamental questions of neuroscience:

How thought and consciousness emerge from the brain and its basic units, the neurons?
Where is memory stored?
If areas of the brain are specialised, how come sometimes they can be substituted by others?

Methods:

Electroenchphalograph (EEG) invented in 1929 (Hans Berger).
Functional MRI since ~2000s
Single-neuron measurements
Nerves-machine interfaces



Neuroscience

Hodgkin and **Huxley** demonstrated experimentally (also studying the giant axon of the squid), that the action potential (”spike”) is the consequence of special Na^+ and K^+ conductances that change with the membrane potential. Their studies lead to the Hodgkin-Huxley mathematical model of the neuron.

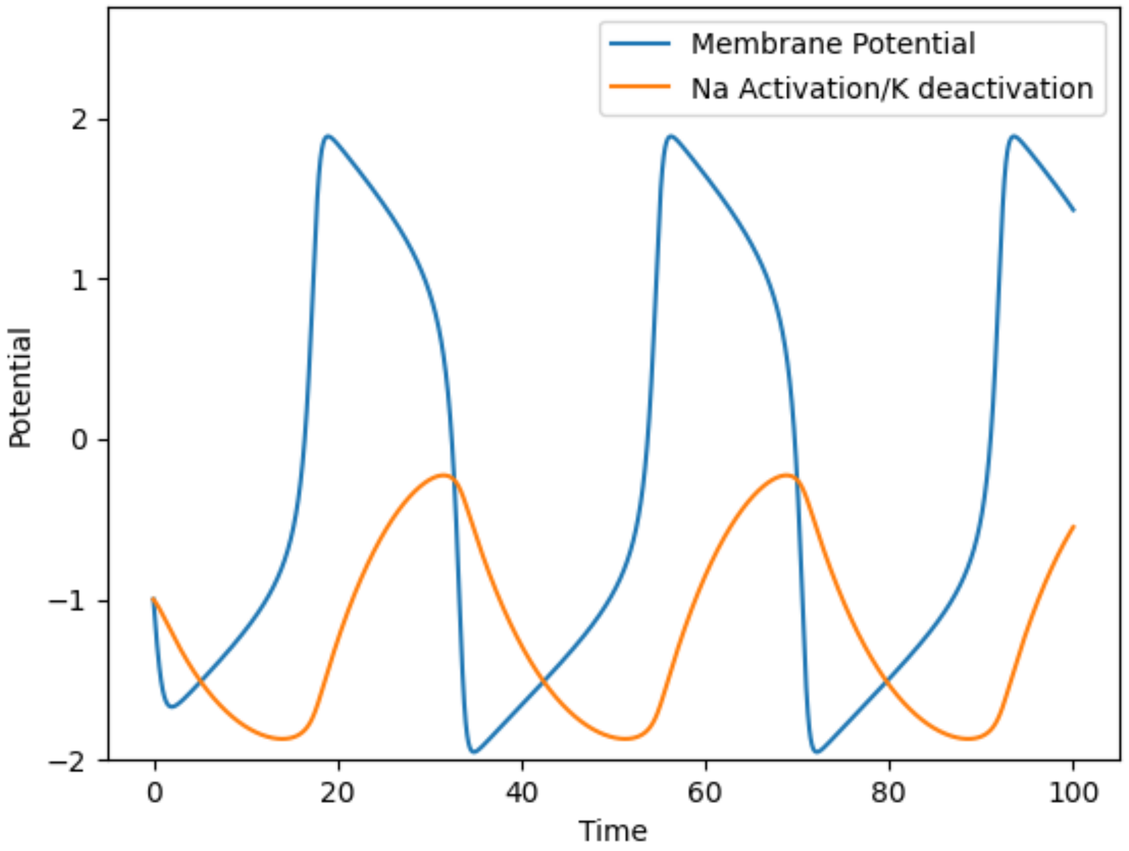
A simplified 2D-model: the FitzHugh–Nagumo model

$$\begin{cases} \dot{v} = v - \dfrac{v^3}{3} - w + RI_{\text{ext}} \\ \tau\dot{w} = v + a - bw. \end{cases}$$

Membrane voltage

Na ion activation channel
K ion deactivation channel

External stimulus: if a threshold is exceeded, the system displays a spiking behaviour



Psychology

How humans (and animals?) think and behave?

Beginning 900s:

Behaviorism
Rejection of introspection in favour of external (more scientific?) observation.
Focus on stimolus->response.

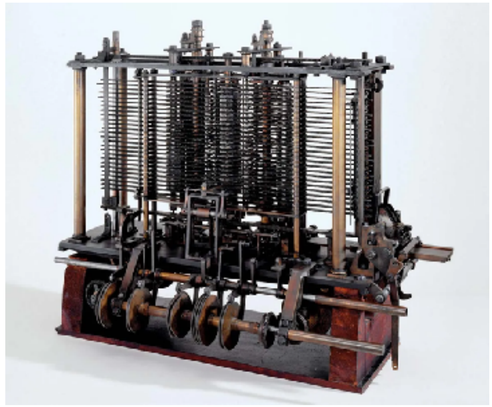
Cognitive psychology
Brain as information-processing device

Beginning 1950s:

Cognitive science
Computer modelling

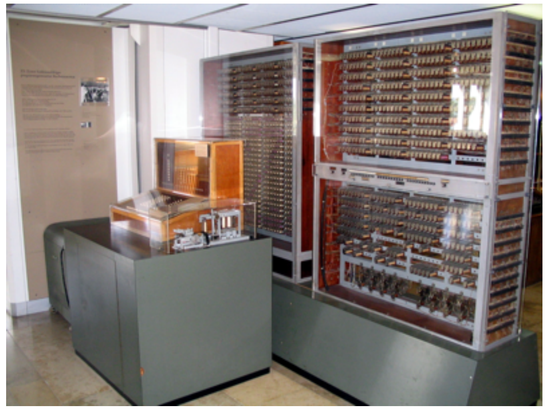
Computer Engineering

Charles Babbage's Difference Engine (1837)



<1Flops

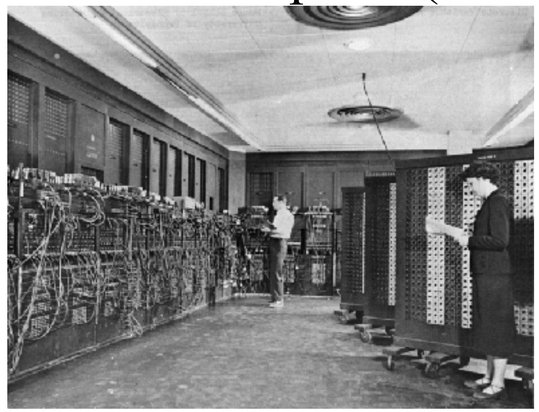
Konrad Zuse's first programmable electro-mechanical computer (1941)



Deutsches Museum reproduction

10Flops

ENIAC: first programmable electronic computer (1945)



(U.S. Army photo, c. 1947–1955)

5kFlops

Today's (2024) supercomputer

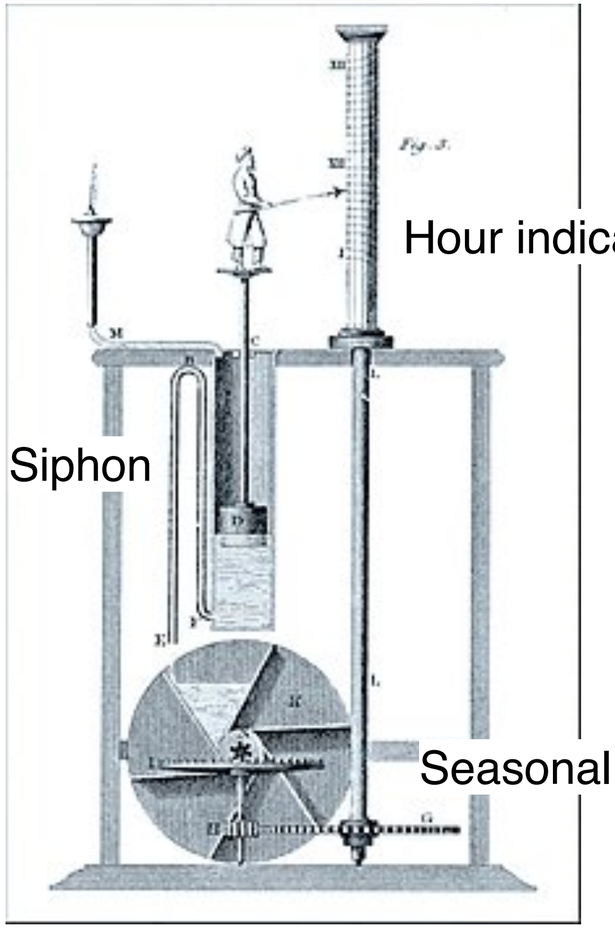
Frontier - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD

8,699,904 cores
1,679.82 PFlops = 1.7 EP!

Control Theory ("Cybernetics")

Central concept: self-regulation

The first known self-regulatory machine:
the **water clock** of Ctesibius of Alexandria (285–222 BC)



Steam-engine governor of James Watt (1736-1819)

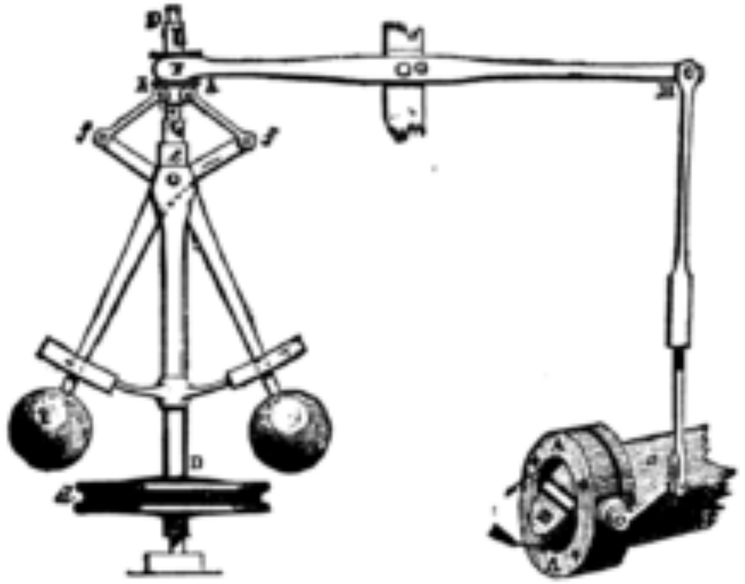


FIG. 4.—*Governor and Throttle-Valve.*

Control Theory ("Cybernetics")

Founder of the field: Norbert Wiener (1894-1964).

- Main work: the *Cybernetics* book (1948)
- Introduced the idea of "intelligent" machines which self-regulate with feedback loops

Today: Control theory is a well developed mathematical theory. The focus is on systems trying to minimise a cost function over time. Also stochastic functions are considered.

The theory can, like AI, be viewed as describing "intelligent agents".
What is the difference?

- Control theory: deeply rooted in the use of tools like linear algebra and calculus ,good for describing systems with continuous, fixed variables.
- AI: broader field which considers also other kinds of variables and mathematical tools.

Linguistics

Relevant field for AI: language is a main human trait.
When learning a language, humans can produce sentences never heard before.

This means that we learn "structures", not specific word sequences?
Noam Chomsky (1928-): Syntactic Structures (1957).

Differences between:
- Syntax: the structure of a sentence
- Semantics: the meaning of a sentence

Colorless green ideas sleep furiously!
Noam Chomsky

Today:
- Computational linguistics
- Natural Language Processing
- Knowledge Representation

Some more definitions/food for thought

Strong AI thesis:
Every thought is an algorithm and conscience too. A computer can completely simulate thoughts.

Weak AI thesis:
Conscience emerges from the physical working of the brain. A computer can simulate every action of the brain with a corresponding algorithm, but the execution of algorithms cannot make conscience emerge.

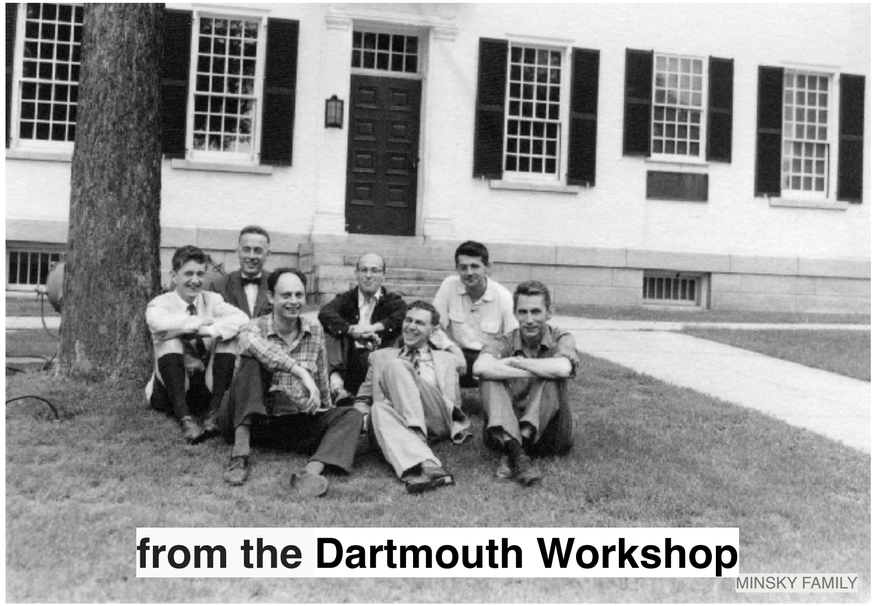
Uncomputability Thesis:
Conscience is an attribute of the brain, but it is not computable (cannot be translated into an algorithm).

Spiritualist Thesis:
Conscience is not an attribute of the brain, but belongs to a "soul": it is not computable nor explainable with science.

X Thesis ….
Y Thesis ….

Short History of AI

1940s: W. McCulloch and W. Pitts: first computer model of a neuron
D. Hebb: “hebbian” learning rule for neurons.
1950: M. Minsky, D. Edmonds: first neural computer (3000 vacuum tubes)
1955: Dartmouth Workshop: first AI workshop
1958: LISP programming language (used in AI)
1950s: some progress on “problem solvers”
1962: The “perceptron” as connected neurons
1969: The perceptron is only a linear separator: NNs dead?
1970-80: A new approach: Expert systems
1986: NNs are back: more layers, back propagation
1990s: Probabilistic reasoning
1996-7: Deep Blue wins against Kasparov at Chess
2000s: Big data (—> training)
2010s: Deep learning



Oliver Selfridge, Nathaniel Rochester, Marvin Minsky, John McCarthy, Ray Solomonoff, Claude Shannon, Peter Milner.

Summary

What is AI?

- Is its aim:
  - -To imitate humans as best as possible or…
  - -…to be able to choose the “right” (best?) decision/action?

-Limits: computational intractability

-Mind and brain coincide? Is it possible to simulate the brain mechanically?

-Curiously:

- Computers had great success in replicating what in the past was thought to be the sign
- of high rationality and intelligence: playing chess, doing calculations, memorising data.
- On the contrary, it does not look so easy to replicate actions which are “simple” for
- humans, e.g. vision.

-Nowadays AI is booming again: faster hardware, more data (internet!), new algorithms.