

# Brief Background of AI

Saumya Jetley

October 4, 2022



## **Intelligence that is implemented artificially**

- through machinery besides the biological machinery
- through machinery besides the human brain

# What is Intelligence?



*The ability to solve problems correctly and efficiently*

- play the rubik's cube
- optimise a path through cities
- drive a car between these cities
- placate a frustrated employee
- teach a student of art

# Characteristics of Intelligence?



Characteristics of intelligence must then necessarily include the ability to:

1. reason from first principles
2. communicate (effective use of language)
3. create (envision new realities)
4. adapt (learn and evolve)



What are we imitating?

Human Thought	Rational Thought
Human Actions	Rational Actions



## 1. Human thinking, uncovered by:

- mapping human thoughts (introspection)
- mapping human actions (psychology)
- mapping human brain (brain imaging)

As the outputs for given inputs match, the machine can be said to be modelling some of brain's mechanisms.



## 1. Human thinking, uncovered by:

- mapping human thoughts (introspection)
- mapping human actions (psychology)
- mapping human brain (brain imaging)

As the outputs for given inputs match, the machine can be said to be modelling some of brain's mechanisms.

(Are algorithms that perform as well as humans, a good model of human performance?)



## 2. Human actions, tested using:

### ■ Turing Test<sup>1</sup>:

- A human agent interacting with an AI agent over a voice call should not be able to tell if it is another human or an AI agent.  
(Natural language communication, knowledge representation, knowledge reasoning, machine learning.)

### ■ Total Turing Test:

- A human agent should be able to send video signals and pass things across the counter to verify the AI agents perceptual abilities.  
(Computer vision, robotic control for movement and manipulation.)

---

<sup>1</sup>Computing Machinery and Intelligence by A. M. Turing, 1950





3. Rational thoughts, includes the study of the *Laws of Thought* i.e. the laws that govern rational or irrefutable thinking.<sup>2</sup>

- **Aristotle and Syllogisms**

- The structure of argument where conclusion is always correct given the premise is correct.
- *All men are mortal, Socrates is man – > Socrates is a mortal.*

- **Logicist Tradition**

- Representing world as a set of facts and relations between them
- Reasoning within the framework to generate new facts
- Challenge: Facts cannot be known with 100% certainty; reasoning could take a long time without meta-logic

---

<sup>2</sup>The Logical Theory Machine by A. Newell and H. A. Simon, 1956



## 4. Rational action:

- Operate autonomously, sense the environment, create, adapt and pursue goals but in what way?
- In a way as to maximise expected gain!
- **Tests and approaches become means to that end**
  - Knowledge representation and reasoning enable rational action
  - Natural language communication enables rational action
  - Accurate perception enables rational action

An agent may act with **Limited Rationality** when there is not enough time.



Asimov's notion of rationality:

- **The first law** is that a robot shall not harm a human, or by inaction allow a human to come to harm.
- **The second law** is that a robot shall obey any instruction given to it by a human, and
- **The third law** is that a robot shall avoid actions or situations that could cause it to come to harm itself.

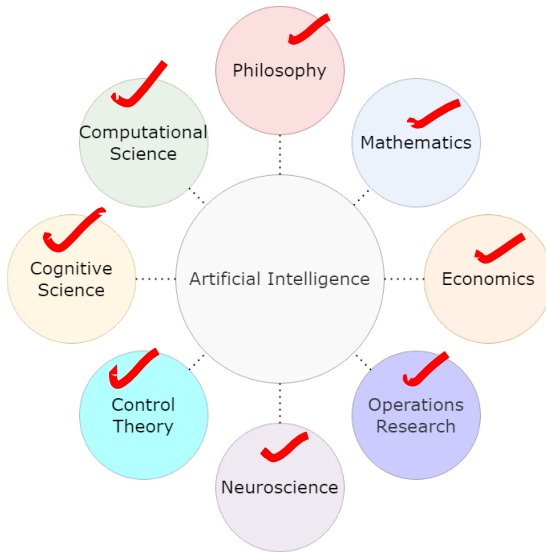


Asimov's notion of rationality:

- **The first law** is that a robot shall not harm a human, or by inaction allow a human to come to harm.
- **The second law** is that a robot shall obey any instruction given to it by a human, and
- **The third law** is that a robot shall avoid actions or situations that could cause it to come to harm itself.

Q: Is this rationality well defined?

# AI as a confluence of multiple fields



**Figure:** Emergence of AI



- **Duality of mind and brain**

*Rene Descartes: There is a part of human mind outside of nature, exempt from physical laws.*

- **Empiricism**

*John Locke: Nothing is in the understanding, which was not first in the senses.*

- **Induction**

*David Hume: General rules are acquired by exposure to repeated association between their elements.*

- **Utilitarianism**

*John Stuart Mill: The idea of rational decision criteria pervades all spheres of human activity.*



- **Logic**

*George Boole and the theory of propositional logic*

*Gottlob Frege and first order logic of objects and relations.*

- **Decidability**

- **Algorithms and Computability**

*Contributions of Persia and one al-Khowarazmi*

*Alan Turing and the Halting Problem*

- **Tractability**

*Cook and Karp show the existence of NP-completeness in combinatorial search and reasoning problems*

- **Probability**



- **Preferred outcomes**

*Adam Smith: Economies are made of individuals trying to maximise their personal well-being.*

- **Utility**

*Leon Walras: Mathematisation of the notion of preferred outcomes - utility of actions.*

- **Decision Theory**

*Making decision under uncertainty*

- **Markov Decision Processes**

*Formalising sequential decision making in uncertainty*

- *Herbert Simon: Models based on **Satisficing** - approximate decision making - give a better characterisation of human behaviour*





Analytical methods to improve decision making.

Maximising payoffs over a sequence of actions:

- Installation of radars by the British
- Logistics planning during Gulf War!
  - departure, routes, arrival of military units

\*Note the contributions of Search and Optimisation algorithms\*



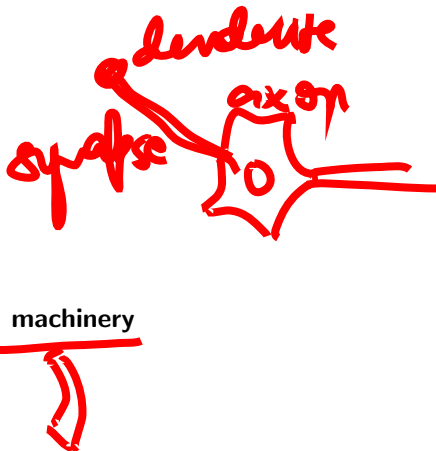
- Specialisation of human brain

- Network of neurons

- Mapping the brain  
*EEG and fMRI*

- Comparison with computational machinery

- Profusion of storage
- Profusion of interconnections





- **Homeostatis** through feedback loops
- **Objective function** that must be maximised



## ■ Psychology

*Introspective: relate mental processes while performing perceptual or associative tasks*

*Behavioral: observe behaviour (actions or responses) whilst considering only objectively measurable percepts (stimulus)*

## ■ Cognitive Science

*Kenneth Craik's 3 steps of a knowledge-based agent - stimulus to internal representation - cognitive processes to create new internal representations - retranslated into action*

*Modelling human cognition using machines*



Some key concepts:

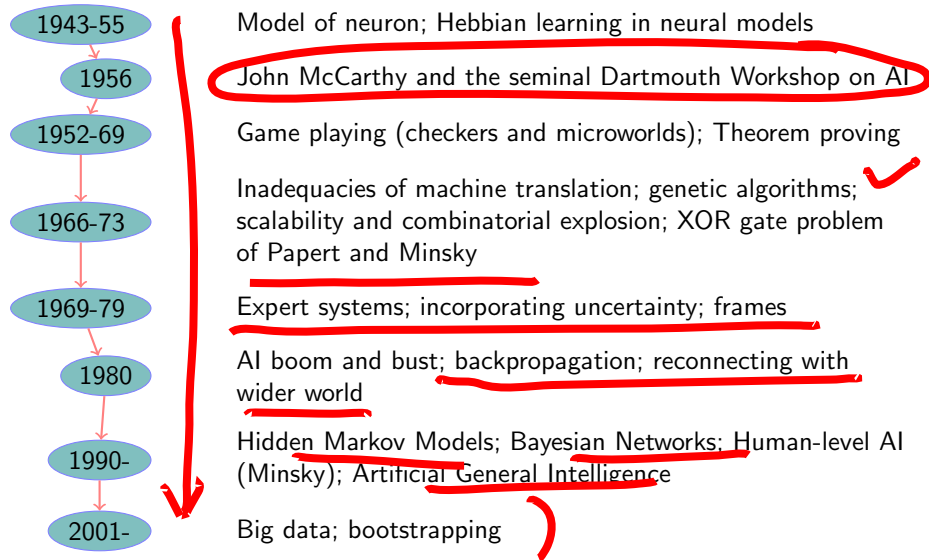
- Modularisation
- Recursion
- Parallelisation

Areas of application:

- Biological
- Linguistics

*computational*

# Evolution timeline





# Overview

## 1 Introduction

- What is AI?
- What is Intelligence?
- Characteristics of Intelligence?
- Humans vs. Rationality
- Asimov's rationality

## 2 AI as an intersection

- Philosophy
- Mathematics
- Economics
- Operations Research
- Neuroscience
- Control theory
- Cognitive Science
- Computational sciences

## 3 Evolution of AI



- Artificial Intelligence - A Modern Approach by Stuart Russell and Peter Norvig, 2010 (Chapter 1)
- Real world challenges for Artificial General Intelligence by Koray Kavukcuoglu, 2021