# Artificial Intelligence[1]

## Introduction

[1]The slides have been prepared using the textbook material available on the web, and the slides of the previous editions of the course by Prof. Luigia Carlucci Aiello

# Summary

◇ What is AI?

◇ A short history

◇ The state of the art

# What is AI?

EU vision:

Artificial intelligence (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, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications).

# What is AI?

| | |
|---|---|
| "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem-solving, learning . . ." (Bellman 1978) | "The study of mental faculties through the use of computational models" (Charniak McDermott 1985) |
| "The study of how to make computers do things at which, at the moment, people are better" (Rich Knight 1991) | "AI ... is concerned with intelligent behaviour in artifacts" (Nilsson, 1998) |

# What is AI?

| | |
|---|---|
| **Systems that think like humans** | **Systems that think rationally** |
| **Systems that act like humans** | **Systems that act rationally** |

# Acting humanly: The Turing test

Turing (1950) "Computing machinery and intelligence":

$\diamondsuit$ "Can machines think?" $\longrightarrow$ "Can machines behave intelligently?"

$\diamondsuit$ Operational test for intelligent behavior: the Imitation Game

$\diamondsuit$ Predicted that by 2000, a machine might have a 30% chance of fooling a lay person for 5 minutes

$\diamondsuit$ Anticipated all major arguments against AI in following 50 years

$\diamondsuit$ Suggested major components of AI: knowledge, reasoning, language understanding, learning

Problem: Turing test is not reproducible, constructive, or amenable to mathematical analysis

# Thinking humanly: Cognitive Science

1960s "cognitive revolution": information-processing psychology replaced prevailing orthodoxy of behaviorism

Requires scientific theories of internal activities of the brain
– What level of abstraction? "Knowledge" or "circuits"?
– How to validate? Requires
1) Predicting and testing behavior of human subjects (top-down)
2) Direct identification from neurological data (bottom-up)

Both approaches (roughly Cognitive Science and Cognitive NeuroScience) are now distinct.

# Thinking rationally: Laws of Thought

Normative (or prescriptive) rather than descriptive

Aristotle: what are correct arguments/thought processes?

Several Greek schools developed various forms of logic:
notation and rules of derivation for thoughts;

Direct line through mathematics and philosophy to modern AI

Problems:
1) Formalize knowledge is difficult
2) How far can we get with computation of inferences?

# Acting rationally

Rational behaviour: doing the right thing

The right thing: that which is expected to maximize goal achievement, given the available information

Does not necessarily involve thinking—e.g., blinking reflex—but thinking should be in the service of rational action

Aristotle (Nicomachean Ethics):

*Every art and every inquiry, and similarly every action and pursuit, is thought to aim at some good*

## Rational agents

An agent is an entity that perceives and acts

This course is about designing rational agents

Abstractly, an agent is a function from percept histories to actions:

$$f : \mathcal{P}^* \rightarrow \mathcal{A}$$

For any given class of environments and tasks, we seek the agent (or class of agents) with the best performance

Caveat (1): *computational limitations make perfect rationality unachievable* $\rightarrow$ design best program for given resources

Caveat (2): Herbert Simon' satisficing behaviour: applies to AI?

Caveat (3) RN4: objective is NOT fixed/known by AI systems

# More mundane views

Intelligent = not stupid

where not is defined as negation as failure (to prove)

AI aims at solving difficult problems. When a problem is solved, it is no longer AI

AI is typically characterized a set of techniques.

# AI and other disciplines

| | |
|---|---|
| Philosophy | logic, methods of reasoning<br>mind as physical system<br>foundations of learning, language, rationality |
| Mathematics | formal representation and proof<br>algorithms, computation, (un)decidability,<br>(in)tractability, probability |
| Psychology | adaptation<br>phenomena of perception and motor control<br>experimental techniques (psychophysics, etc.) |
| Economics | formal theory of rational decisions |
| Linguistics | knowledge representation<br>grammar |
| Neuroscience | plastic physical substrate for mental activity |
| Control theory | homeostatic systems, stability<br>simple optimal agent designs |

# History of AI

1931

The Austrian Kurt Goedel shows that in first-order predicate logic all true statements are derivable. In higher order logics, on the other hand, there are true statements that are unprovable.

1937

Alan Turing points out the limits of intelligent machines with the halting problem.

1943

McCulloch and Pitts model neural networks and make the connection to propositional logic.

1950

Alan Turing defines machine intelligence with the Turing test and writes about learning machines and genetic algorithms

1951

Marvin Minsky develops a neural network machine. With 3000 vacuum tubes he simulates 40 neurons

# History of AI

1955

Arthur Samuel (IBM) builds a learning chess program that plays better than its developer

1956

McCarthy organizes a conference in Dartmouth College. Here the name Artificial Intelligence was first introduced

Newell and Simon of Carnegie Mellon University (CMU) present the Logic Theorist, the first symbol-processing computer program

1958

McCarthy invents at the Massachusettes Institute of Technology (MIT) the high-level language LISP. He writes programs that are capable of modifying themselves

1959

Gelernter (IBM) builds the Geometry Theorem Prover

1961

The General Problem Solver (GPS) by Newell and Simon imitates human thought

# History of AI

1963

McCarthy founds the AI Lab at Stanford University

1965

Robinson invents the resolution calculus for predicate logic

1966

Weizenbaum's program Eliza carries out dialogue with people in natural language

1969

Minsky and Papert show in their book Perceptrons that the perceptron, a very simple neural network, can only represent linear functions

1972

French scientist Alain Colmerauer invents the logic programming language PROLOG

British physician de Dombal develops an expert system for diagnosis of acute abdominal pain. It goes unnoticed in the mainstream AI community of the time

# History of AI

1976
Shortliffe and Buchanan develop MYCIN, an expert system for diagnosis of infectious diseases, which is capable of dealing with uncertainty
1981
Japan begins, at great expense, the Fifth Generation Project with the goal of building a powerful PROLOG machine
1982
R1, the expert system for configuring computers, saves Digital Equipment Corporation 40 million dollars per year
1986
Renaissance of neural networks through, among others, Rumelhart, Hinton and Sejnowski. The system Nettalk learns to read texts aloud
1990
Pearl, Cheeseman, Whittaker, Spiegelhalter bring probability theory into AI with Bayesian networks
Multi-agent systems become popular

# The last 20 years of AI

1997 Deep Blue wins against Kasparov
2005 First DARPA Challenge on Autonomous CAR
2011 IBM Watson defeats Jeopardy champions
2016 Deepmind wins against GO champions
2017 Autonomous taxi
2018-19 Super human AI for multi player poker

# The state of the art: Competitions/Challenges

1997 First RoboCup in Nagoya
2020 24th RoboCup in Bordeaux (France)

- DARPA Grand Challenges
- First Smart City Robotic Challenge
- Everywhere (ICAPS, ... Angry bird)

# State of the art: key progress

Big data and Deep Learning made a breakthrough

- speech understanding
- image classification
- autonomous cars

# State of the art: keys for success

Digital technologies

- computing power;
- massive amounts of data;
- sensor and actuators

Communication technologies

- Internet-5G
- Web

# Computational power for AI

| | Supercomputer | Personal Computer | Human Brain |
|---|---|---|---|
| Computational units | $10^4$ CPUs, $10^{12}$ transistors | 4 CPUs, $10^9$ transistors | $10^{11}$ neurons |
| Storage units | $10^{14}$ bits RAM | $10^{11}$ bits RAM | $10^{11}$ neurons |
| | $10^{15}$ bits disk | $10^{13}$ bits disk | $10^{14}$ synapses |
| Cycle time | $10^{-9}$ sec | $10^{-9}$ sec | $10^{-3}$ sec |
| Operations/sec | $10^{15}$ | $10^{10}$ | $10^{17}$ |
| Memory updates/sec | $10^{14}$ | $10^{10}$ | $10^{14}$ |

| | Supercomputer | Personal Computer | Human Brain |
|---|---|---|---|
| Computational units | $10^6$ GPUs + CPUs | 8 CPU cores | $10^6$ columns |
| | $10^{15}$ transistors | $10^{10}$ transistors | $10^{11}$ neurons |
| Storage units | $10^{16}$ bytes RAM | $10^{10}$ bytes RAM | $10^{11}$ neurons |
| | $10^{17}$ bytes disk | $10^{12}$ bytes disk | $10^{14}$ synapses |
| Cycle time | $10^{-9}$ sec | $10^{-9}$ sec | $10^{-3}$ sec |
| Operations/sec | $10^{18}$ | $10^{10}$ | $10^{17}$ |

# AI main areas (from the Turing test)

- Knowledge Representation–Automated Reasoning
- Machine Learning
- NLP and Speech understanding

- Robotics
- Vision

# Application domains

AI for AgriFood
AI for Cybersecurity
AI for Cultural Heritage
AI for Education
AI for Finance, Commerce and Legal Issues
AI for Health and Medicine
AI for Industrial Automation
AI for Media and Entertainment
AI for the Public Administration
AI for Space
AI for Transportation and Smart Cities ...

# State of the art challenges

One example, Winograd's schema:

> The city councilmen refused the demonstrators a permit because they feared/advocated violence

Hard to generalize
Hard to get training data
Biases in the training data
Lack of transparency / (Explainable AI)

integrate **model-free** and **model-based** approaches

# Philosophical questions

- Can a machine **think** ?
- Can a machine have an **intelligent behaviour** ?
- Can a machine have **consciousness** ?
- Can a machine have **self-consciousness** ?
- Can a machine have **emotions, ...** ?

# Societal issues

- Advantages and disadvantages of **automation**
- Impact on **human intelligence**?
- **Privacy**
- **Legal Issues**
- **Ethics**

see:
"https://ec.europa.eu/digital-single-market/en/artificial-intelligence"

# Ethical principles from the EU Expert Group

1. Human agency and oversight
2. Technical robustness and safety
3. Privacy and Data Governance
4. Transparency
5. Diversity, non-discrimination and fairness
6. Societal and environmental well-being
7. Accountability

# Asimov's laws of robotics

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm;
2. a robot must obey the orders given it by human beings, except when such orders would conflict with the previous law;
3. a robot must protect its own existence as long as such protection does not conflict with the previous two laws.

# Oren Etzioni's laws of AI

1. An AI system must be subject to the full gamut of laws that apply to its human operator.
2. An AI system must clearly disclose that it is not human.
3. An AI system cannot retain or disclose confidential information without explicit approval from the source of that information.

# Clarke's laws

1. When a distinguished but elderly scientist states that something is possible, he is almost certainly right. When he states that something is impossible, he is very probably wrong.
2. The only way of discovering the limits of the possible is to venture a little way past them into the impossible.
3. Any sufficiently advanced technology is indistinguishable from magic.