AILAB
Artificial Intelligence Laboratory

Artificial Intelligence

Prof. Federico Bergenti
Artificial Intelligence Laboratory
Università degli Studi di Parma
www.ailab.unipr.it/bergenti

Class Details

- Class schedule
  - Wednesday 8:30am–10:30am, room G
  - Thursday 8:30am–10:30am, room G

- Classes are recorded
  - The course page on Elly (elly2024.didattica.unipr.it) mentions the code to register to the team and access the recordings
  - Only *formally* enrolled students can access the classes and the recordings
  - *It is strictly prohibited to download the recordings*

Communications with the Teacher

- Few simple rules to communicate with the teacher
  - Communications and announcements from the teacher for the students are posted via Elly
  - Meetings with the teacher are requested via e-mail
  - Before requesting a meeting, send an email to the teacher describing the reasons for the meeting request
    - Clearly describe the problems with sufficient details
    - Attach the exercises that originated the meeting request

federico.bergenti@unipr.it

Exam Details (I)

- Exam schedule
  - 3 sessions in January–February
  - 3 sessions in June–July
  - 1 session in September

- Exam sessions are composed of
  - A written exam and an oral exam
  - The oral exam follows a successful written exam
  - The oral exam is upon explicit request from the student *or from the teacher*
  - The oral exam is part of the process, and it is not supposed to only improve the grade of the written exam
  - Written and oral exams must be passed in the same session
  - Written and oral exams cover the whole course program

Exam Details (II)

- Exam enrollment
  - Enrollment is strictly needed to participate to an exam
  - Enrollment is performed using Esse3

unipr.esse3.cineca.it

  - Enrollment is possible only within the period associated with each session
  - Clearing the enrollment for an exam is possible only within the period associated with each session
- Students are not supposed to enroll to all sessions

Didactic Materials

- Classes are the official reference for the course
  - The slides used for the classes will be available via Elly throughout the academic year (after classes)
  - Class recordings will be available via Elly throughout the academic year
- Additional materials will be provided via Elly
  - To expand and complement slides
  - To allow for deeper investigations on selected topics
- Further reading
  - S. Russell, P. Norvig. *Artificial Intelligence: A Modern Approach*, 4th edition, Pearson, 2020

What is Artificial Intelligence?

"In the 1980s, Minsky and Good had shown how neural networks could be generated automatically—self replicated—in accordance with any arbitrary learning program. Artificial brains could be grown by a process strikingly analogous to the development of a human brain."

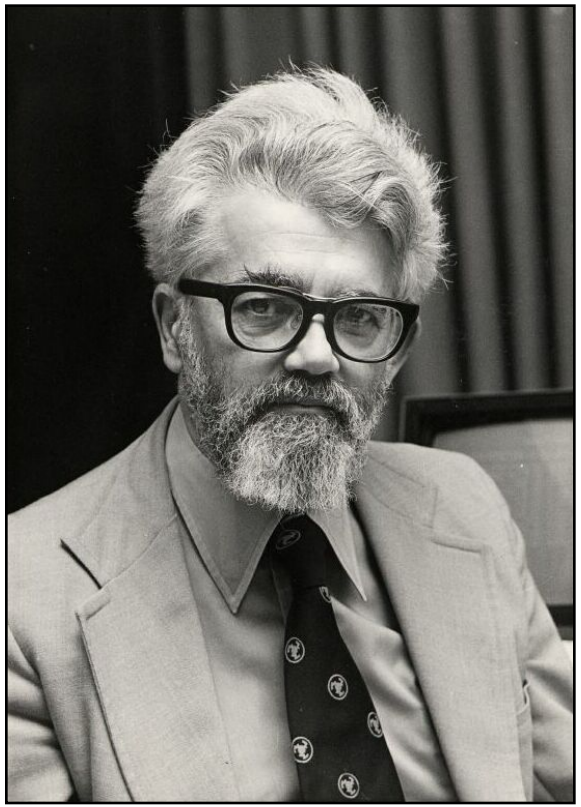
Arthur C. Clarke, *2001: A Space Odyssey*

Artificial Intelligence

- An ambitious project started together with Computer Science
- The goal of the project was to design and develop
  - Machines with intelligent behaviors
  - Machines that could effectively interact with the real world (robots)
- Recently, Artificial Intelligence (AI) often targets
  - Machines that can solve complex problems
  - Machines that exhibit rational behaviors
  - Machines that can dialog with humans in a natural language
  - Machines that can generate creative contents
  - Machines that interact with complex and dynamic worlds (the Internet, the Web, ...)

The Dartmouth Meeting

- Prof. John McCarthy (1927-2011) invents the name *artificial intelligence* in 1955
  - In a letter to propose a meeting at the Dartmouth College to be held in the summer of 1956
- The Dartmouth meeting hosted discussions on the problems that AI have *not yet* solved
  - What is intelligence?
  - What is rationality?
  - Can machine think?
    - If so, how?
    - If not, why not?
  - …



A Characterization of AI

- From a published interview with John McCarthy
  - *Q: What is artificial intelligence?*
  - A: 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.
  - *Q: Yes, but what is intelligence?*
  - A: Intelligence is the computational part of the ability to achieve goals in the world. Varying kinds and degrees of intelligence occur in people […]

Other Characterizations

- AI is meant to deal with complex problems that can be (easily) solved by humans and other living beings, but that are not normally described as algorithms
  - E.g., understand natural languages, invent proofs of theorems, play chess or Go
- Several problems that were targeted by AI in the past are no longer considered relevant
  - E.g., compile a Fortran program (~1955), compute the symbolic indefinite integral of a function (~1965), find a known structure in an image (~1970)

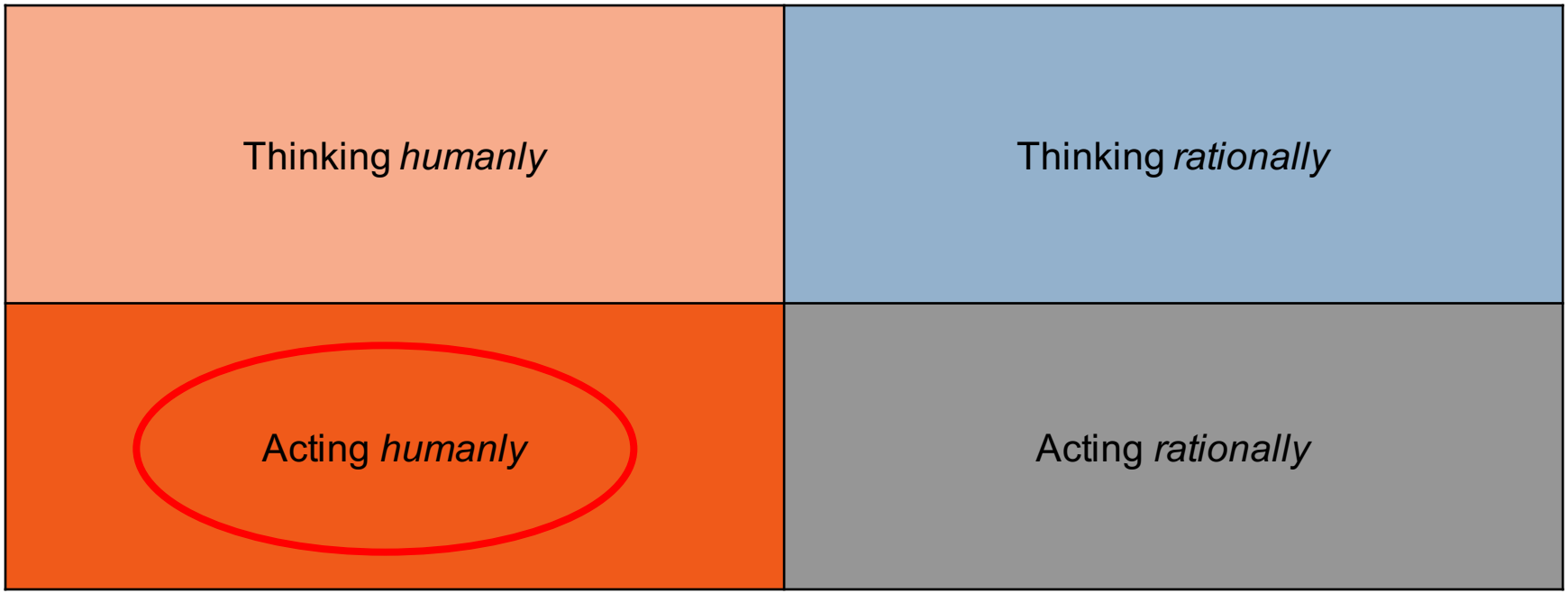
What is Easy? What is Hard?

- Several activities that are based on *cognitive capabilities* and that are normally associated with intelligence have already been automatized algorithmically, sot they are no longer relevant for AI
- Several activities that are easily and routinely performed by animals have not yet been automatized algorithmically
  - Walk without bumping into obstacles
  - Fuse information from senses (touch, sight, smell, …)
  - Effectively work together without a centralized control (like bees and ants)

Long Term Goals of AI

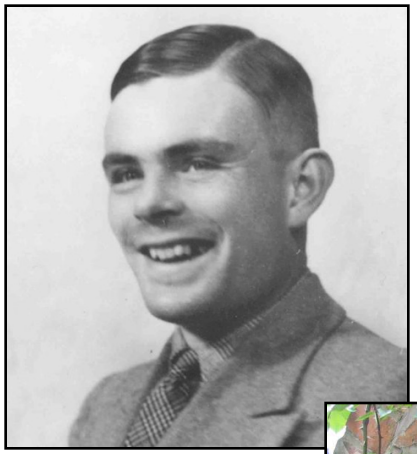
- AI can be strong
  - A computer executing an appropriate program can exhibit an intelligence that is indistinguishable from human intelligence
  - The empiricist philosophers Thomas Hobbes stated *"By ratiocination, I mean computation* […] *Ratiocination, therefore, is the same with addition and subtraction."* (Elements of Philosophy, 1656)
- AI can be weak
  - A computer will never have sufficient resources to exhibit an intelligence that is indistinguishable from human intelligence
  - A computer will never be so complex
    - Computers will simulate selected processes of the human mind

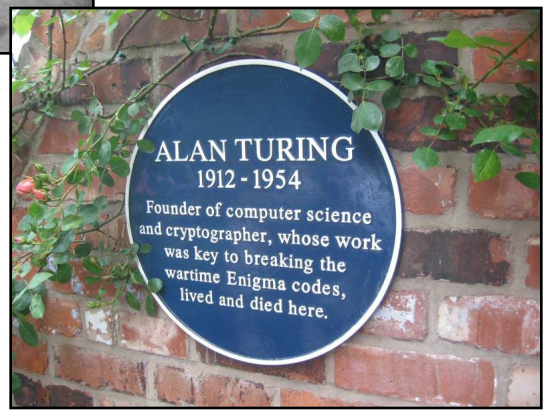
Four Viewpoints



Turing Test (I)

- What are the essential characteristics of intelligent behavior?
- Dr. Alan Turing (1912-1954) proposed an empirical test in 1950
- The Turing test is still relevant to understand the goals of AI





Turing Test (II)

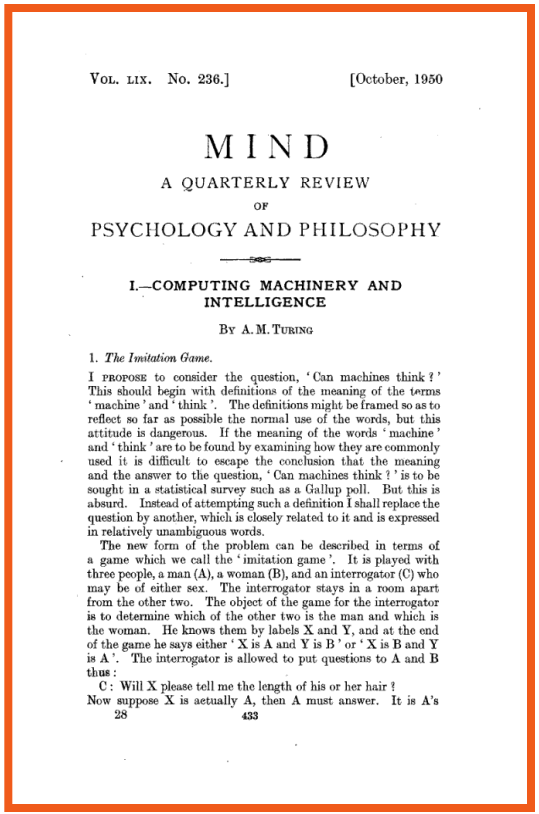
- The article Computing Machinery and Intelligence (Mind, 1950) starts with a section entitled

  The Imitation Game

- The section starts as follows

  I propose to consider the question,
  "Can machines think?"

- The article describes a test to assess if a machine can be considered intelligent
  - It provides an empirical definition of intelligence



VOL. LIX. No. 236.] [October, 1950

MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

BY A. M. TURING

1. *The Imitation Game.*

I PROPOSE to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, 'Can machines think?' is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the 'imitation game'. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either 'X is A and Y is B' or 'X is B and Y is A'. The interrogator is allowed to put questions to A and B thus:

C: Will X please tell me the length of his or her hair?

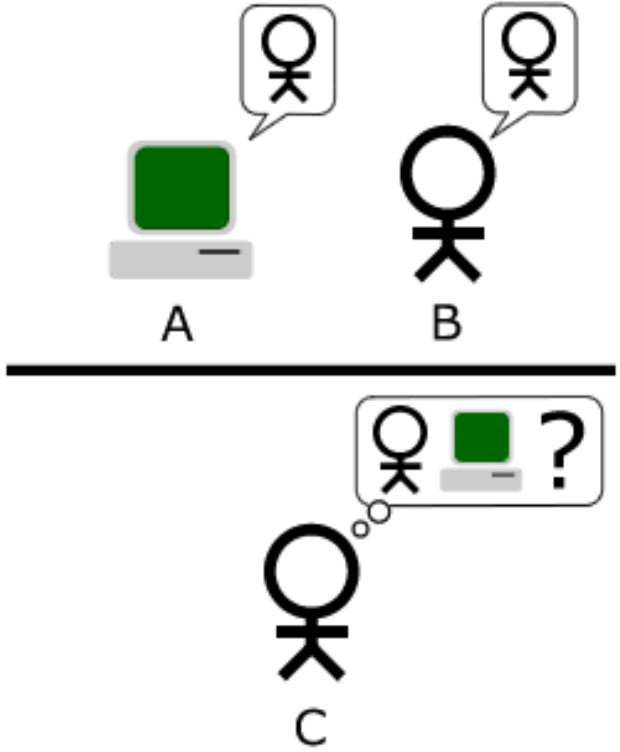
Now suppose X is actually A, then A must answer. It is A's

28 433

Turing Test (III)

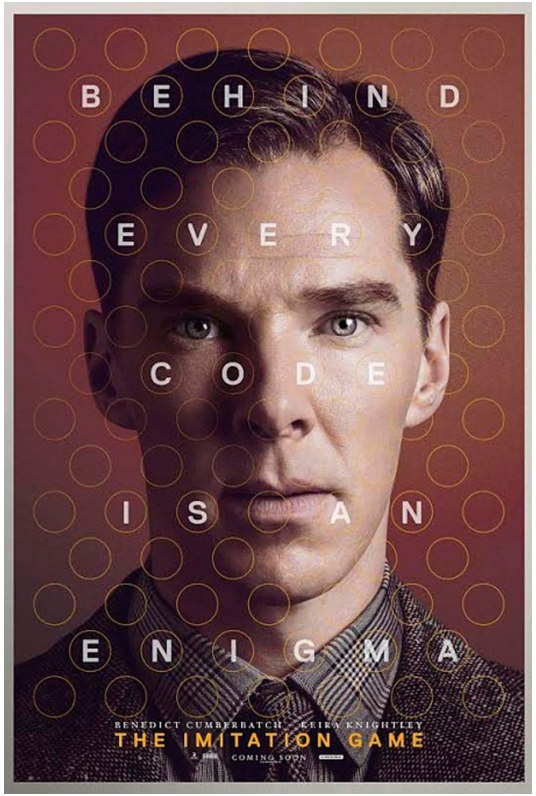
- *A* is a machine designed and developed to play chess
  - E.g., IBM *Deep Blue*
- *B* and *C* are two chess players (not newbies, not champions)
- *C* is not in the same room where *A* and *B* are
- *C* does not known if the current opponent is either *A* (the machine) or *B* (the human)

- *A* can be considered *as intelligent as B* if *C* cannot decide which is the current opponent



Acting Humanly

- The Turing test provides an empirical definition for the term *intelligent behavior*
  - In the original test, Turing did not talk about chess players, which inherently require rationality
- The Turing test advocates an anthropocentric approach
  - The goal is to act as a person
  - Intelligent machines play *imitation games*



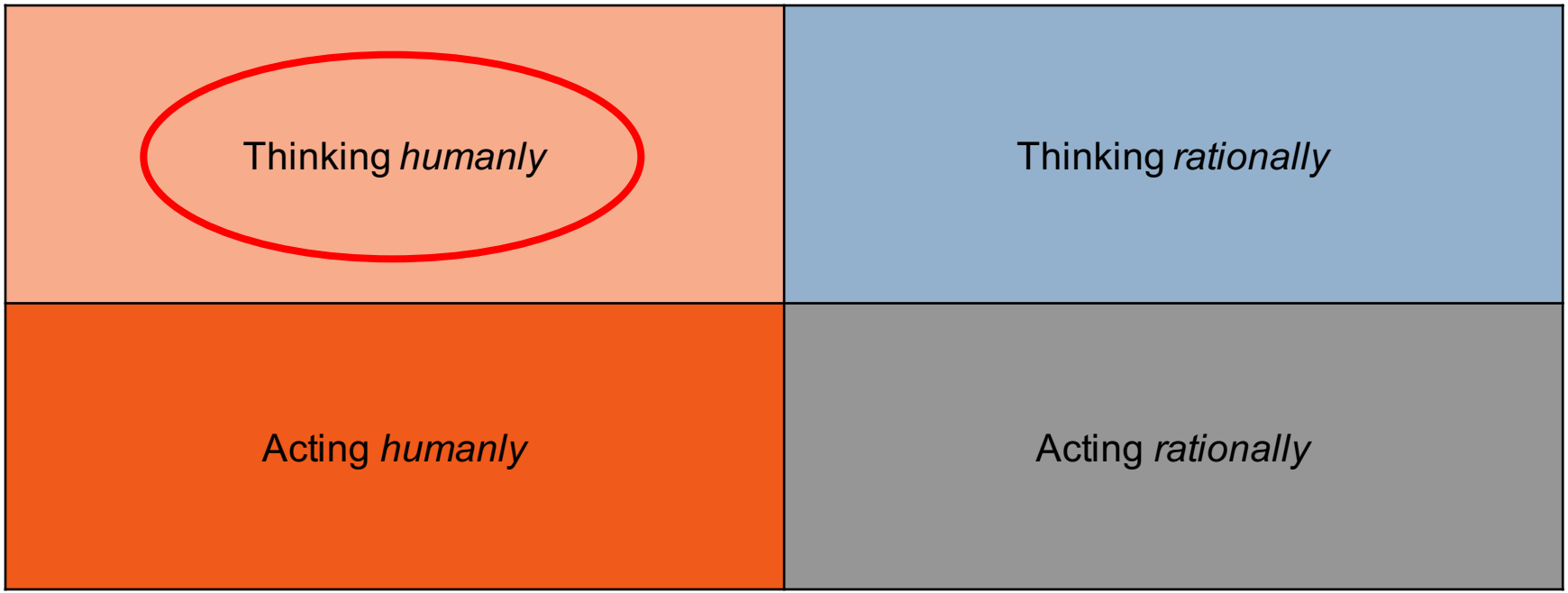
Eliza (J. Weizenbaum, 1966)

- The Turing test has been approached several times by several computer programs
- The computer program Eliza behaves as a Rogerian psychotherapist

- Eliza simply
  - Identifies relevant parts of input sentences
  - Rephrases input sentences to continue the conversation

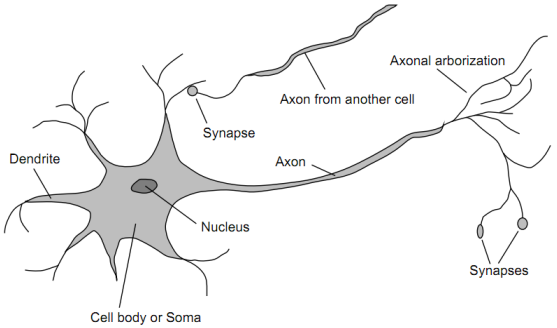
```
> Hello, I am Eliza.
* Hi
> How are you today...What would you like to discuss?
* I have problems at school
> Can you elaborate on that?
* I don't pass exams
> Do you wish to be able to pass exams?
* Yes, sure
> I see.
* Math exams are too difficult for me
> Come, come, elucidate your thoughts.
* I don't see a meaning in math
> Do you wish to be able to see a meaning in math?
```

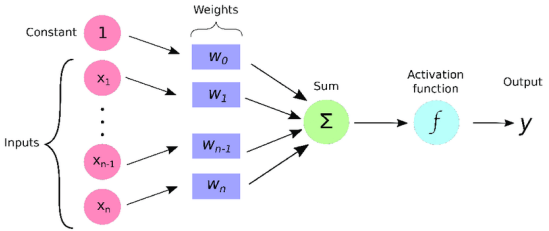
Four Viewpoints



Thinking Humanly

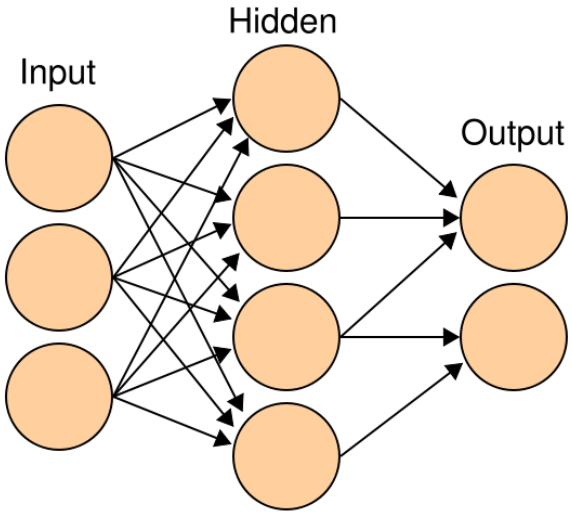
- A possible approach to imitate human behavior is based on the possibility to simulate the organ that originates thoughts
  - The brain is the organ where thoughts originate
- Using this approach, the goal is to build *electronic brains* to simulate human brains
  - The simulation is at the cell level
  - The simulation includes simulated neurons, axons, soma, …





Neural Networks (I)

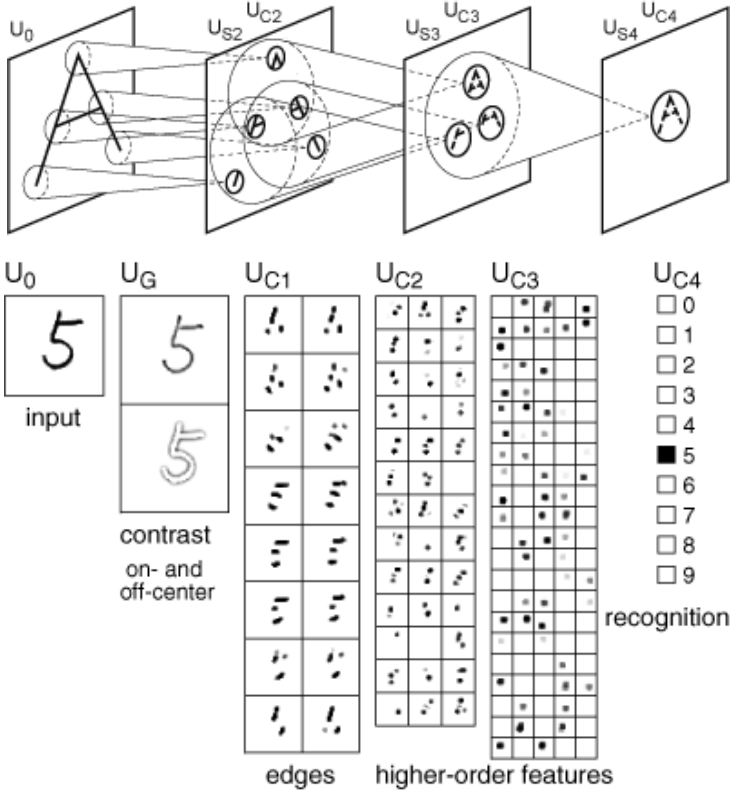
- Each *unit* simulates a neuron
- Units are connected in a *network*
  - That receives *percepts* from *sensors*
  - That provides *stimuli* to *actuators*
- The network is trained to *learn* the desired behavior
  - Normally, it is too complex to be explicitly programmed
  - Useful neural networks have billions of neurons



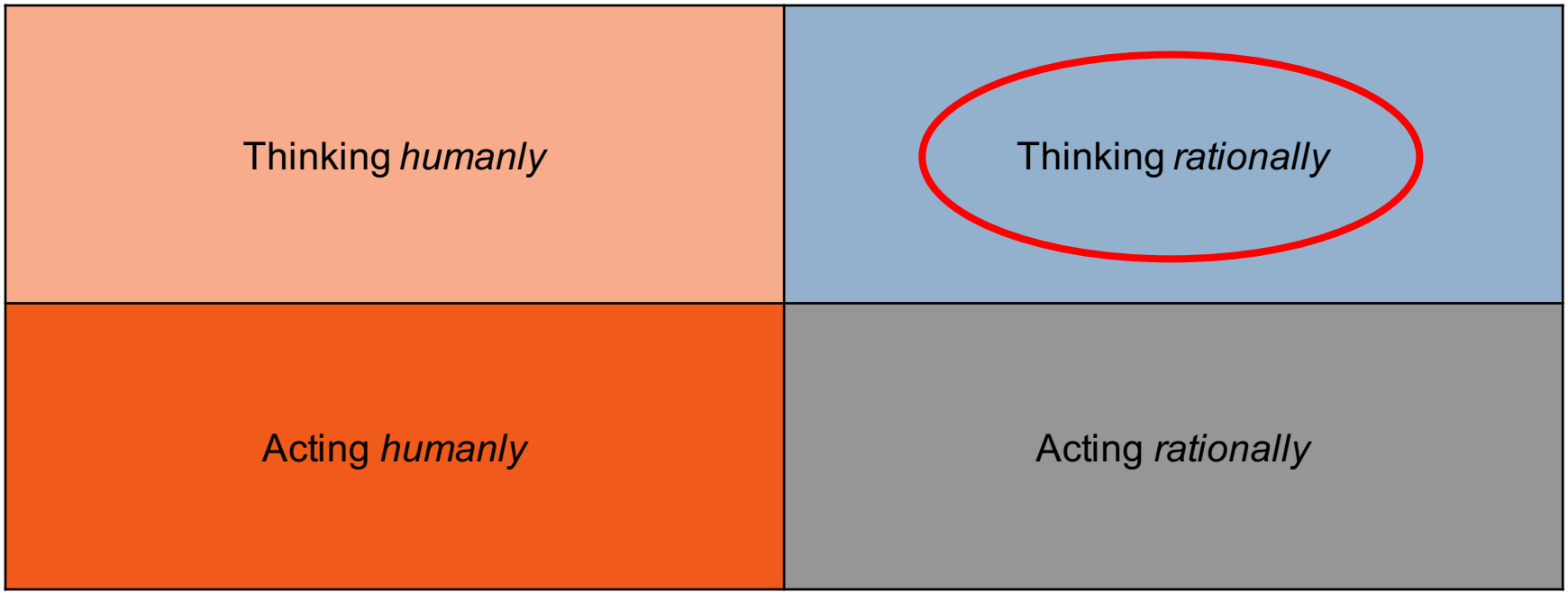
Neural Networks (II)

- *Neocognitron* (K. Fukushima, 1980)
  - Neural network trained to recognize handwritten characters

- It is an example of a
  - *Convolutional Neural Network* (*CNN*) because each neuron receives input from a local subset of the precedent neurons
  - *Deep network* (i.e., a network with several layers) that leverages deep learning



Four Viewpoints



Thinking Rationally

- Human beings are not always rational
  - They are heavily influenced by habits, hopes, unreachable goals, unsolvable problems, …
- Rationality is described by logic languages or equivalent formalisms
  - Assume that $A \rightarrow B$ is considered true
  - B must be considered true as soon as A is known to be true
- Logic formalisms are used for deductions

Expert Systems

- A computer program written in terms of a *knowledge base*, which is a (possibly dynamic) set of *facts* and *inference rules*
  - Facts:

$$mother\_of(ann, bob), sister\_of(claire, ann)$$

  - Inference rules:

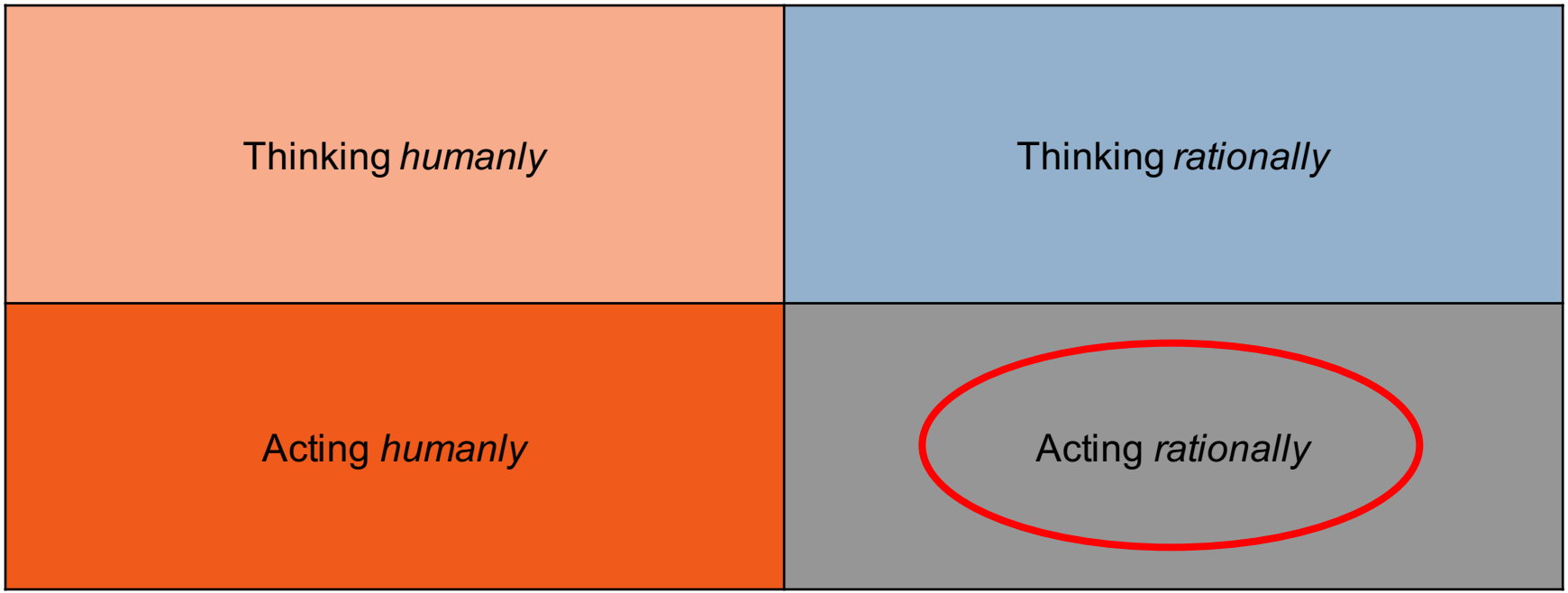
$$mother\_of(X, Y) \wedge sister\_of(Z, X) \rightarrow aunt\_of(Z, Y)$$

- The knowledge base of an expert system can be used to *infer* new facts from known facts

MYCIN (E. Shortliffe, 1970s)

- MYCIN is an expert system for medical diagnosis that
  - Contains more than 600 inference rules
  - Can ask yes/no questions
  - Provides a list of diagnosis and prescriptions
- An extensive experimental campaign measured that MYCIN was correct in 69% of the cases
  - It was correct more often than the doctors that provided the inference rules

Four Viewpoints



Acting Rationally

- The machines that act rationally within an *environment* are known as intelligent agents
  - An agent wants the *world* (agent and environment) in certain states, and it acts *autonomously* to bring the world there
  - An agent is characterized by its behavior and the mechanisms used to obtain its behavior are irrelevant
    - Agents are often based on logic languages
    - Agents often use deduction to decide what to do (i.e., to decide which action to perform on the environment)
    - Agents are not requested to imitate human behaviors
    - Agents are normally requested to be rational

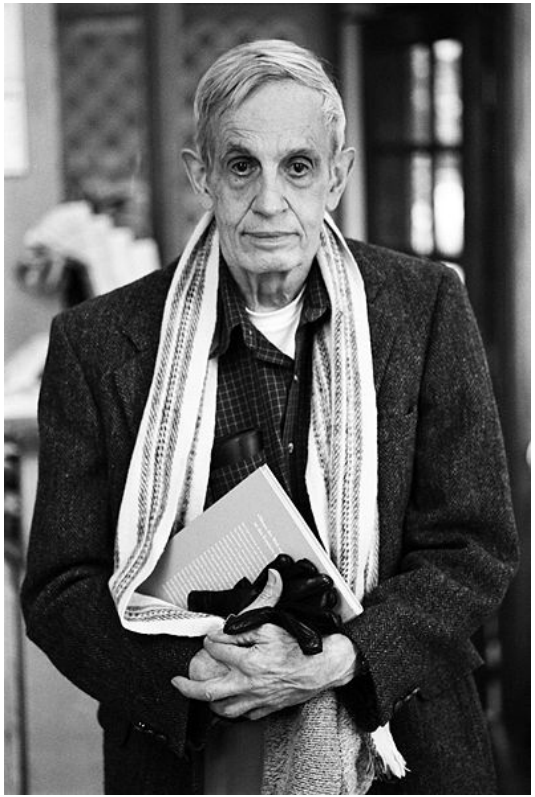
SHRDLU (T. Winograd, 1971)

- The computer program SHRDLU controls an emulated robot arm to interact in the blocks world
- The user can interact with SHRDLU using a natural language (English)
  - SHRDLU associates a meaning with user sentences
  - SHRDLU can ask questions to better understand the user
- SHRDLU has been considered for several years the prototypical AI program

```
* Pick up a big red block.
> OK.
* Find a block which is taller than
the one you are holding and put it
into the box.
> BY "IT", I ASSUME YOU MEAN THE
BLOCK WHICH IS TALLER THAN THE ONE I
AM HOLDING.
> OK.
* What does the box contain?
> THE BLUE PYRAMID AND THE BLUE
BLOCK.
* What is the pyramid supported by?
> THE BOX.
* How many blocks aren't in the box?
> FOUR OF THEM.
* Is at least one of them narrower
than the one which I told you to
pick up?
> YES, THE RED CUBE.
```

Game Theory

- Game theory is used to design and develop intelligent agents, especially when the world is a multi-agent system
  - It was popularized by the works of John Nash (1928-2015)
  - It is studied and used, differently, by AI and Economics
- Game theory is used to reason on rational behaviors by describing rational behaviors in terms of the earned utility



Game Theory and AI

- AI often restricts its scope to games with the following characteristics
  - The game includes only two players that play alternate turns
  - The game has simple and formalized rules
  - The game is played in a completely accessible environment, and therefore, players have *perfect knowledge* of the state of the game
  - The game (often) have strict time constraints
- For example
  - Checkers, chess, … are often considered in AI
  - Poker, bridge, … are not often considered in AI

Real Agents (I)

- *Chinook* won the *Man-Machine Checkers Championship*
- It is still possible to play against Chinook

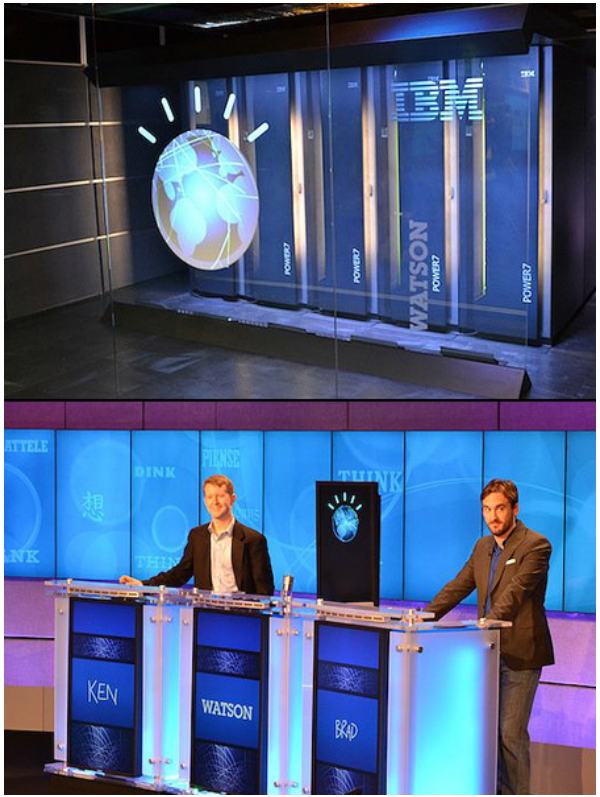
  www.cs.ualberta.ca/~chinook
- IBM *Deep Blue* won against Garry Kasparov in 1996 and received worldwide attention also from general media

Real Agents (II)

- Go was considered unfeasible for computers
  - Chess: *branching factor* > 40, more than 50 turns per game
  - Go: branching factor > 250, more than 350 turns per game
- *AlphaGo* is an AI player for Go
  - In October 2015, AlphaGo won 5-0 against the European (human) champion
- AlphaGo uses neural networks to support a game-theoretic algorithm to decide next move

Real Agents (III)

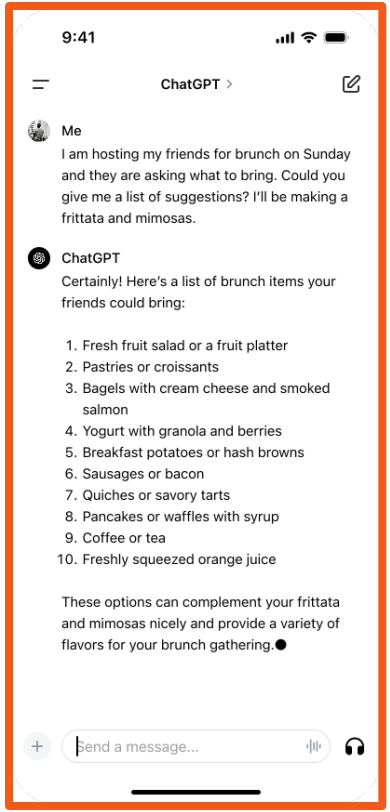
- *Watson* is a computer program that can answer questions expressed in English
  - It stores ~$200 \cdot 10^6$ documents including a dump of Wikipedia
- In 2011, Watson participated to three special episodes of *Jeopardy!* winning against several opponents
  - Watson was not connected to the Internet during the show recordings



Real Agents (IV)

- *GPTs* (*Generative Pre-trained Transformers*) are a family of related *LLMs (Large Language Models)*
- LLMs are deep neural networks trained to complete texts
  - They are trained on a huge amount of texts produced by humans (typically, more than 1T words of documents)
  - They are fine tuned using reinforcement learning feedback from humans
- LLMs are examples of generative AI agents

- Several alternatives to GPTs exist (e.g., LLAMA, Gemini, …)



Artificial Intelligence

Prof. Federico Bergenti
Artificial Intelligence Laboratory
www.ailab.unipr.it/bergenti

www.ailab.unipr.it

Artificial Intelligence Laboratory
Università degli Studi di Parma
ailab@unipr.it