Artificial Intelligence Notes

Marco Natali

INDICE

ELENCO DELLE FIGURE

1 INTRODUCTION

This course will provide an introduction to AI techniques and approach analyzed nowadays and to understand the current state of art we have to provide an Timeline to see progress and discover done during the time, so in figure 1 we will see all important events related with AI.

The major discover happened on 2020 are the following:

GPT3 (GENERATIVE PRE-TRAINED TRANSFORMER): produced by OpenAI in May 2020, is a larger and richer language model consisting in 175 billion machine learning parameters used for automatic text generation, translation, user interface synthesis

DARPA CHALLENGE (ALPHADOGFIGHTS) with simulated F-16 Air Fighters where on $18 - 20$ August 2020 there was the final Event, where AI system was against each other and the winner was a system by Heron system, that was also able to defeated a human expert top gun fighter $5 - 0$.

On [**?**] Andrew NG says that AI will transform many industries, but it's not magic and almost all of AI's recent progress is based on one type of AI, in which some input data (A) is used to quickly generate some simple response (B) $[A \rightarrow B]$.
Also Andrew Ng says that if a typical person can do a mental task with less than one second of thought, we can probably automate it using AI either now or in the near future.
Choosing A and B creatively has already revolutionized many industries, it is poised to revolutionize many more.

ML systems are not (yet?) able to justify in human terms their results, so for some application it is essential the human knowledge to be able to generate explanations, infact some regulations requires the right to an explanation in decision-making, and seek to prevent discrimination based on race, opinions, health, sex and so on, like GPDR.
ML systems learn what's in the data, without understanding what's true or false, real or imaginary, fair or unfair and so it is possible to develop bad/unfair models.

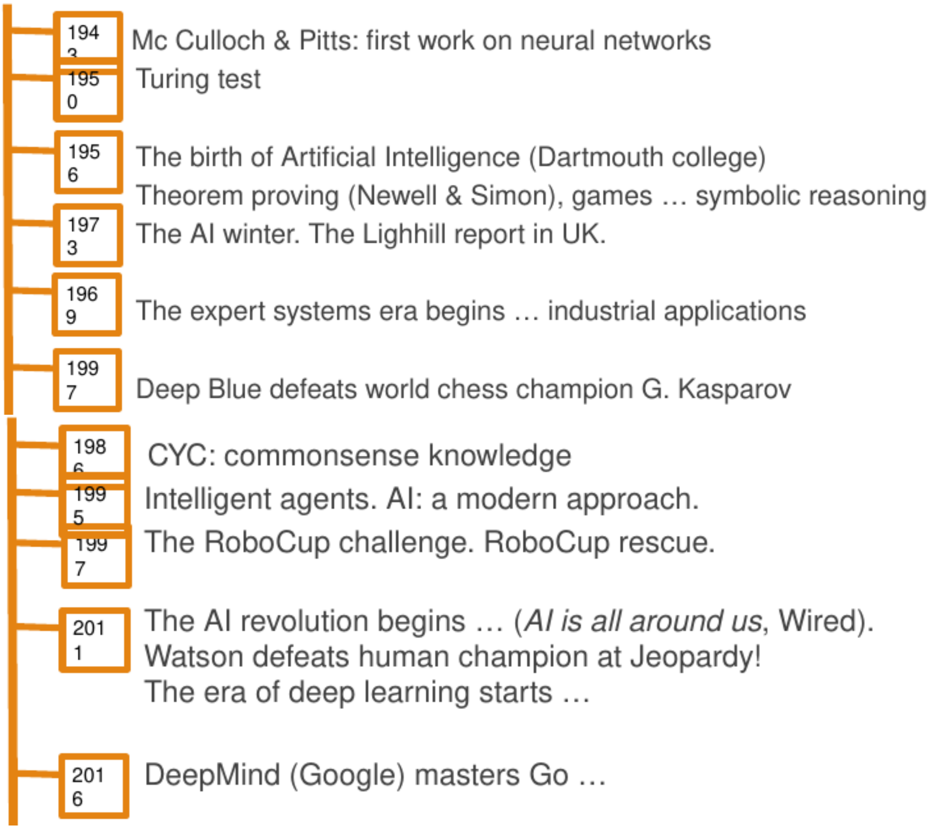
The goal of building AI systems is far from being solved and is still quite challenging in its own. Building complex AI systems requires the combination of several techniques and approaches, not only ML.
One of the most challenging tasks ahead of us is integration of perception and reasoning in AI systems.

AI fundamentals is mostly about "Slow thinking" or "Reasoning" and AI fundamentals has the role, within the AI curriculum, of teaching you about the foundations of a discipline which is now 60 year old.
We will cover different approaches, also some coming of the "Good Old- Fashioned Artificial Intelligence" (GOFAI) or symbolic AI.

Def. Symbolic AI is an high-level "symbolic" (human-readable) representations of problems, the general paradigm of searching for a solution, knowledge representation and reasoning, planning.
Symbolic AI was the dominant paradigm of AI research from the mid 1950s until the late 1980s and central to the building of AI systems is the *Physical symbol systems hypothesis*, formulated by Newell and Simon.

The approach is based on the assumption that many aspects of intelligence can be achieved by the manipulation of symbols (the physical symbol system hypothesis):

Figura 1: AI Timeline evolution


1943 Mc Culloch & Pitts: first work on neural networks
1950 Turing test
1956 The birth of Artificial Intelligence (Dartmouth college)
Theorem proving (Newell & Simon), games … symbolic reasoning
1973 The AI winter. The Lighhill report in UK.
1969 The expert systems era begins … industrial applications
1997 Deep Blue defeats world chess champion G. Kasparov
1986 CYC: commonsense knowledge
1995 Intelligent agents. AI: a modern approach.
1997 The RoboCup challenge. RoboCup rescue.
2011 The AI revolution begins … (AI is all around us, Wired).
Watson defeats human champion at Jeopardy!
The era of deep learning starts …
2016 DeepMind (Google) masters Go …

Def. A physical symbol system has the necessary and sufficient means for general intelligent action

Human thinking is a kind of symbol manipulation system (a symbol system is necessary for intelligence) and machines can be intelligent (a symbol system is sufficient for intelligence).
The hypothesis cannot be proven, we can only collect empirical evidence and observations and experiments on human behavior in tasks requiring intelligence.
We have two different typologies of AI, that was introduced and considered:

STRONG AI: relies on the strong assumption that human intelligence can be reproduced in all its aspects (general A.I.).
It includes adaptivity, learning, consciousness and not only pre-programmed behavior.

WEAK AI: simulation of human-like behavior, without effective thinking/understanding and no claim that it works like human mind; it is the dominant approach today.

A problem of AI is that computer can't have needs, cravings or desires and Abraham Maslow's define a hierarchy of human needs:

1. Biological needs (food, sleep, sex, ...)
2. Safety, protection from environment
3. Love and belonging, friendship
4. Self esteem and respect from others
5. Self-actualization

2 AGENTS

Artificial intelligence, or AI, is the field that studies the synthesis and analysis of computational agents that act intelligently.
An agent is something that acts in an environment/it does something and we are interested in what an agent does, that is, how it acts, so we judge an agent by its actions.

An agent acts *intelligently* when what it does is appropriate given the circumstances and its goals, it is flexible to changing environments and changing goals, it learns from experience and it makes appropriate choices given its perceptual and computational limitations.

Def. A *computational agent* is an agent whose decisions about its actions can be explained in terms of computation.

We have that the central scientific goal of AI is to understand the principles that make intelligent behavior possible in natural or artificial systems, instead the central engineering goal of AI is the design and synthesis of useful, intelligent artefacts, agents, that are useful in many applications.
This is done by the analysis of natural and artificial agents, formulating and testing hypotheses about what it takes to construct intelligent agents and in the end designing, building, and experimenting with computational systems that perform tasks commonly viewed as requiring intelligence.

Artificial Intelligence is not the opposite of real Intelligence, infact intelligence cannot be fake, so if an artificial agent behaves intelligently, it is intelligent and it is only the external behavior that defines intelligence (weak AI).

Artificial intelligence is real intelligence created artificially and we can use different test to estabilish if AI is intelligent: *Turing test* where only external behavior counts and *Winograd schemas* as a test of intelligence, where we asks "The city councilmen refused the demonstrators a permit because they feared violence. Who feared violence?" and "The city councilmen refused the demonstrators a permit because they advocated violence. Who advocated violence?".
These questions are difficult for a machine because it has not the knowledge of context.

The obvious naturally intelligent agent is the human being and the human intelligence comes from three main sources:

1. biology: Humans have evolved into adaptable animals that can survive in various habitats.

2. culture: Culture provides not only language, but also useful tools, useful concepts, and the wisdom that is passed from parents and teachers to children. Language, which is part of culture, provides distinctions in the world that should be noticed for learning.

3. life-long learning (experience): Humans learn throughout their life and accumulate knowledge and skills.

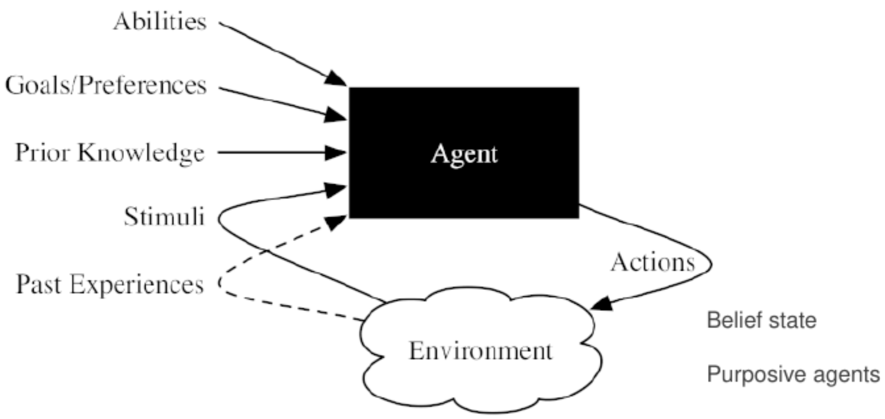
Another form of intelligence is *social intelligence*, the one exhibited by communities and organizations.

Three aspects of computation that must be distinguished:

1. Design time computation, that goes into the design of the agent

2. Offline computation, that the agent can do before acting in the world

Figura 2: Design Process of AI agents



3. Online computation, the computation that is done by the agent as it is acting.

Designing an intelligent agent that can adapt to complex environments and changing goals is a major challenge, so to reach this ultimate goal, two strategies are possible:

1. simplify environments and build complex reasoning systems for these simple environments.
2. build simple agents for natural/complex environments, simplifying the tasks.

The design process of an agents has the following phases, that can be viewed on figure 2:

1. Define the task: specify what needs to be computed
2. Define what constitutes a *solution* and its quality: optimal solution, satisficing solution, approximately optimal solution, probable solution.
3. Choose a formal representation for the task; this means choosing how to represent knowledge for the task and this includes representations suitable for learning.
4. Compute an output
5. Interpret output as a solution

A model of the world is a symbolic representation of the beliefs of the agents about the world, so it is necessarily an abstraction.
More abstract representations are simpler and human-understandable, but they may be not effective enough and low level descriptions are more detailed and accurate but introduce complexity.

Multiple level of abstractions are possible (hierarchical design), but usually two levels are considered:

1. The *knowledge level*: what the agent knows and its goals
2. The *symbol level*: the internal representation and reasoning algorithms

In agent design we consider these aspects, with a summary foundable in figure 3:

MODULARITY: is the extent to which a system can be decomposed into interacting modules and it is a key factor for reducing complexity.
In the modularity dimension, an agent's structure is one of the following:

- *flat* where there is no organizational structure.

- *modular*, which the system is decomposed into interacting modules that can be understood on their own.
- *hierarchical*, which the system is modular, and the modules themselves are decomposed into simpler modules and the agent reasons at multiple levels of abstraction.

PLANNING HORIZON: is how far ahead in time the agent plans and in this dimension an agent is one of the following:

- *Non-planning agent*, that does not look at the future.
- *Finite horizon* planner, where agent looks for a fixed finite number of stages and it is greedy if only looks one time step ahead.
- *Indefinite horizon* planner is an agent that looks ahead some finite, but not predetermined, number of stages.
- *Infinite horizon* planner is an agent that keeps planning forever

REPRESENTATION: concerns on how the state of the world is described and a state of the world specifies the agent's internal state (its belief state) and the environment state.
We have 3 type of representation, from simple to complex:

- *atomic* states, as in problem solving.
- *feature-based* representation: set of propositions that are true or false of the state, properties with a set of possible values.
- Individuals and relations (often called relational representations) and is the Representations at the expressive level of FOL (or contractions).

COMPUTATIONAL LIMIT: an agent must decide on its best action within time constraints or other constraints in computational resources (memory, precision, ...).
The computational limits dimension determines whether an agent has *perfect rationality*, where an agent is able to reasons about the best action without constraints and *bounded rationality*, where an agent decides on the best action that it can find given its computational limitations.
An *anytime algorithm* is an algorithm where the solution quality improves with time and to take into account bounded rationality, an agent must decide whether it should act or reason for longer.

LEARNING: is necessary when the designer does not have a good model and the learning dimension determines whether knowledge is given in advance or knowledge is learned from data or past experience.
Learning typically means finding the best model that fits the data and produces a good predictive model and in this course only modelling formalisms and approaches are dealt, infact all the issues concerned with learning are dealt in the Machine Learning course.

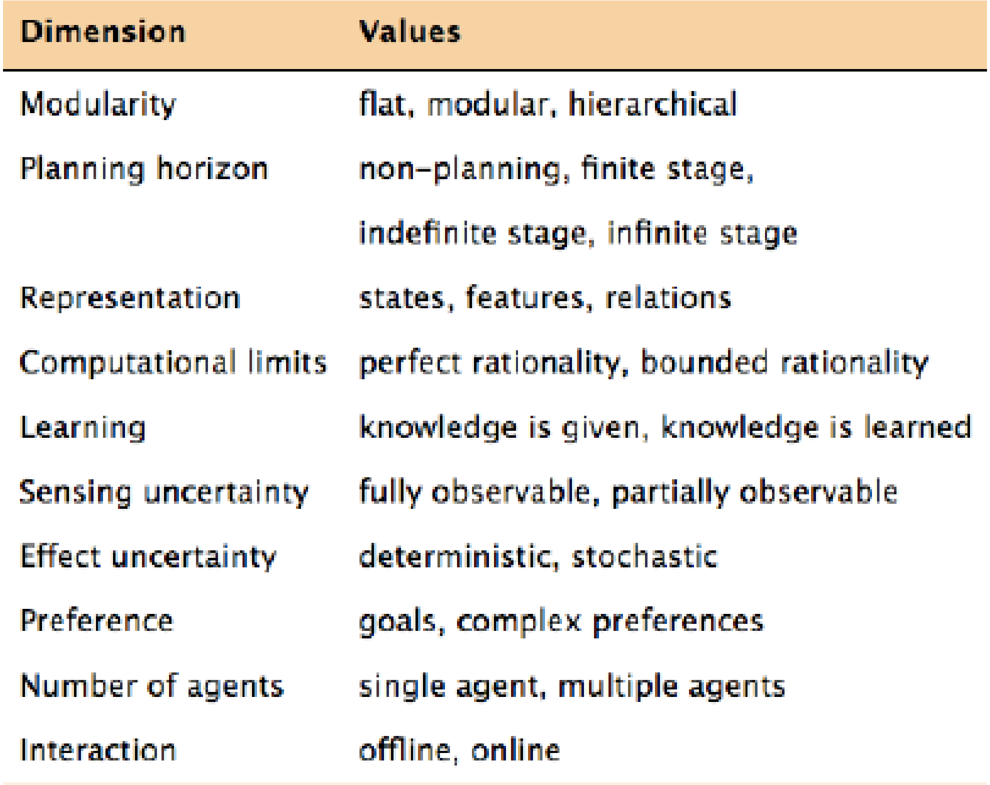
UNCERTAINTY: is divided into two dimensions:

1. uncertainty from sensing/perception (fully observable, partially observable states).
2. uncertainty about the effects of actions (deterministic, stochastic) and when the effect is stochastic, there is only a probability distribution over the resulting states.

PREFERENCE: considers whether the agent has goals or richer preferences:

- A *goal* is either an achievement goal, which is a proposition to be true in some final state, or a maintenance goal, a proposition that must be true in all visited states.

Figura 3: Summary of Agent Design consideration



Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic
Preference	goals, complex preferences
Number of agents	single agent, multiple agents
Interaction	offline, online

- *Complex preferences* involve trade-offs among the desirability of various outcomes, perhaps at different times.
  An *ordinal preference* is where only the ordering of the preferences is important, instead a *cardinal preference* is where the magnitude of the values matters and States are evaluated by utility functions.

NUMBER OF AGENTS: considers whether the agent explicitly considers other agents:

- *Single agent* reasoning means the agent assumes that there are no other agents in the environment or that all other agents are "part of nature", and so are non-purposive.
- *Multiple agent* reasoning means the agent takes the reasoning of other agents into account and this occurs when there are other intelligent agents whose goals or preferences depend, in part, on what the agent does or if the agent must communicate with other agents.

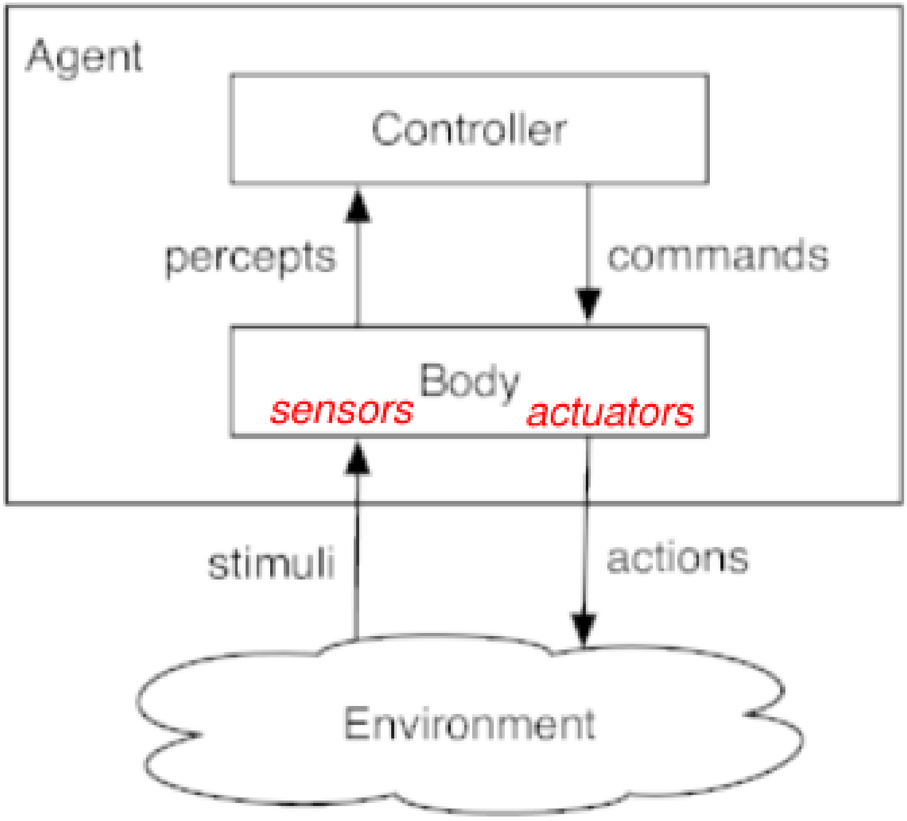
INTERACTION: considers whether the agent does:

- *offline reasoning*, where the agent determines what to do before interacting with the environment.
- *online reasoning*, where the agent must determine what action to do while interacting in the environment, and needs to make timely decisions.

More sophisticated agents reason while acting and this includes long-range strategic reasoning as well as reasoning for reacting in a timely manner to the environment.

Def. An agent is something that interacts with an environment, receiving information through its sensors and acts in the world through actuators (effectors).

Figura 4: Structure of an generic agent system



A robot is an artificial purposive embodied agent and also a computer program is a software agent

An agent is made up of a body and a controller, where the controller receives percepts from the body and sends commands to the body.
A body includes sensors that convert stimuli into percepts and actuators that convert commands into actions. Both sensor and actuators can be uncertain, the controller is the brain of an agent and in the end an agent system includes an agent and the environment in which it acts.
In figure 4 is possible to see the structure of an generic agent system.

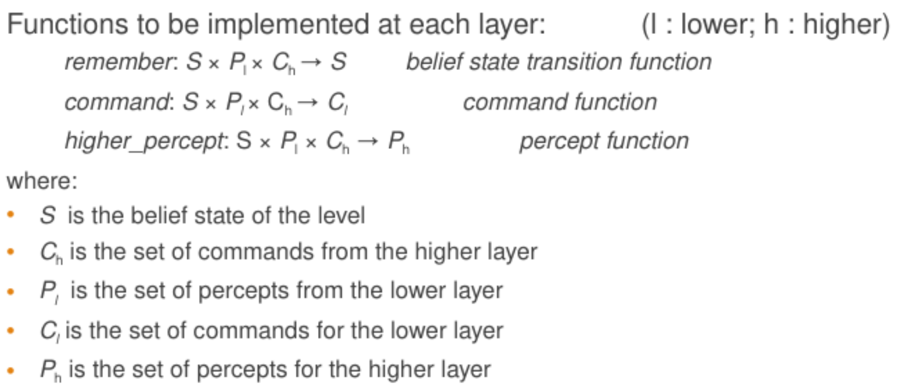
Agents act in time, we have T that is the set of time points, and we assume that T has a start (0) totally ordered, discrete, and each t has a next time $t+1$.
We have an agent history that at time t has percepts up to t and commands up to $t-1$ and we have also *causal transduction* (usually implements by a controller), a function from history to commands, called causal because only previous and current percepts and previous commands can be considered.

However, complete history is usually not available and we have only the memory of it: the memory or belief state of an agent at time t is all the information the agent has remembered from the previous times and the behavior of an agent is described by two functions:

- A *belief state* transition function $S \times P \to S$, where S is the set of belief states and P is the set of percepts.

- A *command* function $S \times P \to C$, where C is the set of commands.

Figura 5: Agent Functions to implement at each layer



Functions to be implemented at each layer: (l : lower; h : higher)

remember: $S \times P_l \times C_h \rightarrow S$ *belief state transition function*

command: $S \times P_l \times C_h \rightarrow C_l$ *command function*

higher_percept: $S \times P_l \times C_h \rightarrow P_h$ *percept function*

where:

- S is the belief state of the level
- C_h is the set of commands from the higher layer
- P_l is the set of percepts from the lower layer
- C_l is the set of commands for the lower layer
- P_h is the set of percepts for the higher layer

The controller implements a command function (an approximation of a causal transduction) and with a single controller it is difficult to reconcile the slow reasoning about complex high-level goals with the fast reaction that an agent needs for lower-level tasks such as avoiding obstacles.
In figure 5 is possible to note the agent functions that should be implemented at each layer of our agent.

high-level reasoning, is often discrete and qualitative • low-level reasoning is often continuous and quantitative A controller that reasons in terms of both discrete and continuous values is called a hybrid system.

The notion of belief state is quite general, most agents need to keep a model of the word and update it while acting and there are 2 extremes:

1. the agent possess a very good predictive model, so it does not need to use perceptions to update the model
2. purely reactive systems do not have a model, and decide only on the basis of perceptions

In the general case the agent uses a combination of prediction and sensing:

- In Bayesian reasoning (under uncertain information) the estimation of the next belief state is called *filtering*.
- In alternative, more complex models of the world can be kept and updated, for example through vision and image processing.

Knowledge of a specific domain may also be represented explicitly and used to decide the action and the *knowledge base* contains general rules and specific/contingent facts in declarative form.
The KB is built offline, built by designers or learned from data and a domain ontology gives meaning to symbols used to represent knowledge; knowledge may be then updated and used to decide actions during operation.