

Artificial Intelligence Fundamentals

Federico Matteoni

A.A. 2021/22

Index

0.1	Introduction	2
-----	------------------------	---

0.1 Introduction

Prof.s: Maria Simi, Vincenzo Lomonaco

AI is taking over the world. Formalizing common sense is a lot more difficult. We can formalize knowledge in very specific and small domains. But is deep learning the final solution to AI? "It will transform many industries, but it's not magic. Almost all of AI's recent progress is based on one type of AI, in which some input is used to quickly generate simple response." (*Andrew Ng*)

This AI can do supervised learning, but requires huge amount of data (tens of thousands of pictures to build a photo tagger, for example). The rule of thumb of Ng is: if a person can do a mental task with less than one second of thought, we can automate it using AI either now or in the near future.

The challenges are:

Software is not a problem, the community is open and the software can be replicated the software can be replicated

Data is exceedingly difficult to get access to. Data is the defensible barrier for many businesses

Talent, because downloading and applying open-source software to your data won't work. AI needs to be customized to context and data, that's why there's a war for the scarce AI talent that can do this work.

Computational resources are also very important.

Deep Learning Is only one approach inside the much wider field of ML and ML is only one approach in the wider field of AI. Book: *Thinking Fast and Slow*, Kahneman. Two systems: system 1 does perceptual tasks, simple computations, system 2 instead does complex computation, recalling from memory. . . this is a distinction in our brains.

Machine Learning Is AI all about machine learning? Possible arguments against ML are:

Explanation and accountability: ML systems are not (yet?) able to justify in human terms their results. For some applications this is essential: knowledge must be meaningful to humans to be able to generate explanations? Some regulations requires the right to an explanation in decision-making, and seek to prevent discrimination based on race, opinions, sex. . . (see GDPR)

ML systems learn what's in the data, **without understanding what's true or false, real or imaginary, fair or unfair**. It is possible to develop unfair, bad models. People are generally more critical about information.

Building AI systems is a goal far from being solved, still quite challenging. Complex AI systems requires the combination of several techniques and approaches, not only ML.

AI Fundamentals Is mostly about reasoning and *slow thinking*. Different approaches, "good old-fashioned artificial intelligence" or "symbolic AI": teaching about the foundations of the discipline, now 60 years old.

Symbolic AI High-level human readable representations of problems, the general paradigm of searching for a solution, knowledge representation and reasoning, planning. Dominant paradigm from the mid 1950s until late 1980s. Central to the building of AI systems is the physical symbol systems hypothesis (PSSH), formulated by Newell and Simon (*Computer Science as Empirical Inquiry: Symbols and Search*)

The approach is based on the assumption that many aspects of intelligence can be achieved by the manipulation of symbols (the PSSH): *a physical symbol system has the necessary and sufficient means for general intelligent action*. Human thinking is a kind of symbol manipulation system (so a symbol system is **necessary** for intelligence), and machine can be intelligent (a symbol system is **sufficient** for intelligence). This cannot be prove, we can only collect empirical evidence: observation and experiments on human behavior in tasks requiring intelligence, and solving tasks of increasing complexity.

Strong and Weak AI The Chinese room argument, by John Searle, introduced the following distinction: strong ai relies on the *strong* assumption that human intelligence can be reproduced in all its aspects (general AI), including adaptivity, learning, consciousness. . . , while weak AI is the simulation of human-like behavior, without effective thinking or understanding, no claim that it works like the human mind. Dominant approach today, fragmented AI. One strong argument against strong AI is the lack of needs by the systems: biological need, safety, relationships, self esteem, self-actualization (Maslow's hierarchy of needs).

What stands in the way of all-powerful AI is not a lack of smarts: it's that computers can't have needs, cravings or desires.

AI is the enterprise of building intelligent computational agents

Agents An agent is something that acts in an environment. We are interested in what an agent does, that is how it acts. We judge an agent by its action. 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, learns from experience, makes appropriate choices given its perceptual and computational limitations.

Computational agent is an agent whose decisions about its actions can be explained in terms of computation and implemented on a physical device.

Scientific Goal: understand the principles that make intelligent behavior possible in natural or artificial systems

Engineering Goal: design and synthesis of useful, intelligent artifacts, agents that are useful in many applications

Artificial Intelligence Artificial intelligence is not the opposite of real intelligence. Intelligence cannot be *fake*: in an artificial agent behaves intelligently, it is intelligent. It is only the external behavior that defines intelligence, according to the **Turing Test** (weak AI). So **artificial intelligence is real intelligence created artificially**.

More updated test: Winograd schemas.

Human intelligence: biology (surviving various habitats), culture (language, tools, concepts, wisdom passed from parents and teachers to children) and life-long learning experience (learning throughout life). Another form is social intelligence, exhibited by communities and organizations.

So agents are situated in environments, inputs are abilities, goals, prior knowledge, stimuli and past experiences, and outputs actions which affect the environment.

