Artificial Intelligence: a concise introduction

Vicenç Torra

School of Informatics
University of Skövde
SE-54128 Skövde, Sweden
e-mail: vtorra@his.se

Chapter 1. Introduction

The number of applications that use artificial intelligence (AI) is large, some of them are used without the users knowing the technology inside. Others appear in the media because they are shocking and/or are considered examples of technological advance.

In this document we give a short summary of the history of artificial intelligence, starting with some of the applications that are or were innovative. Then, we discuss why artificial intelligence is studied and the different definitions there exist for this field. We will also review the main characteristics of artificial intelligence systems, and the main areas of the field. We finish with a few (additional) elements for discussion related to AI.

1 AI into practice: some innovative applications

Methods and techniques developed in the field of AI since 1955 and the espectacular gain on computational power have permited the development of applications that are in use and are successful in their domains. We present below some examples of these applications, classified by topics.

- **Games**. For a long time, optimal decision making in games has been considered as a paradigm of intelligence. A few games are now *solved* (i.e., the outcome can be predicted from any position), others are expected to be solved in the near future. For the most rellevant games there are now programs that beat the best human players.
  - **Checkers**. The computer program Chinook [43] developed by J. Schaeffer and his team from 1989 won the Man-Machine World Championship in 1994. The program includes a database with openings of the best players and another database with end-games with eight pieces or fewer. The same team proved in 2007 that when both players play perfectly, none of them can win. The difficulty of this proof

was on the fact that there are $5 \cdot 10^{20}$ possible positions. [29] explains how Checkers is solved. [28] describes the history of Chinook and the team that developed it.

  - **Chess**. For a long time, chess was the difficult game to be studied. This *problem* was closed in May 1997 when Deep Blue [46], a program developed by IBM, won G. Kasparov in New York. The program used dedicated hardware, databases for openings and end-games. The number of positions in chess it is estimated around $10^{46.5}$.
  - **Go**. Once Deep Blue won Kasparov, research on games turned to Go. This was an open problem until 2016. Only the introduction of Monte Carlo tree search (around 2009) in computer go programs improved their performance significantly. Until then, Go was considered an extremely difficult problem. Finally in March 2016, the computer program AlphaGo beat Lee Sedol in a series of five matches. AlphaGo won the first three matches, then Lee won the fourth, and AlphaGo won the final match. The number of positions in go is estimated as 10^{170} (for a 19x19 board).

- **Intelligent vehicles**. There are different types of vehicles and different degrees of autonomy achieved.
  - **Sendai subway train**. This was the first automatic train in the world, implemented using fuzzy logic by the group led by Seiji Yasunobu (Hitachi). The system was deployed in 1987. Several unmaned trains run today.
  - **Autonomous cars**. Stanley was the winner of the "2005 DARPA Grand Challenge". It was an autonomous car that was able to complete the 212 km off-road course in the Mojave desert (USA) in 6 hours and 54 minutes. Since 2005 several initiatives exist to further develop autonomous cars as e.g. MadeInGermany and Google Self-Driving Car. Tesla autopilot (to be used only in limited-access highways) is currently available[1].
  - **Unmanned Aerial Vehicles (UVA) or drones**. Most are remotely controled, although autonomous ones are also studied and developed.

- **Robotics**. This area has been developed with different goals in mind: automation in industry, militar applications, space exploration. Research include robots of different types, including humanoid robots.
  - **Robotic pets**. Paro (development in 1993-2001) and Aibo (1999) are two examples of them. The first one is a therapeutic robot. Its goal is to stimulate patients with cognitive disorders.

[1] Self-driving car accident on 7 May 2016. The accident at New York Times: http://www.nytimes.com/interactive/2016/07/01/business/inside-tesla-accident.html?_r=0

  - **Humanoid robots**. First biped walking robots with humanoid form were developed around 1995. Honda P3 was probably the first one (1997) and ASIMO (Advanced Step in Innovative MObility, 2000) was designed by Honda after the completion of the P- series. iCub, developed from 2006 is an open source humanoid robot with the size of a 2.5 year old child.
  - **Exploratory and rescue robots**. Two robots were sent to Mars in 2004. They were Spirit and Opportunity. There is a rescue robot competition (the rescue robot league) to foster the research in this area. These competitions were initiated after the RoboCup Soccer leagues (for robots playing soccer).

- **Theorem provers and knowledge-based systems**. The goal of the first is to prove new results (theorems) from a set of premises (axioms). The goal of the later is to make recommendations or decisions based on knowledge represented within the system and the information gathered by the system itself. Expert systems in medicine are an example of knowledge-based systems. They are to find the illness of the patient using domain knowledge and information from the patient.
  - **Theorem provers**. Logic theorist developed by Newell and Simon in 1955 was able to prove theorems from the book Principia Mathematica by Russell and Whitehead. AM (by Lenat in 1977) built theories automatically in mathematics, established new concepts (as e.g. prime numbers) and conjectures, and tried to prove them. It was one of the first discovery systems. There is however some discussion on the real significance of the results of AM. EQP proved in 1996 the Robbins problem: are all Robbins algebras Boolean?. This problem was presented in the thirties by H. Robbins after E. V. Huntington introduced in 1933 the Robbins algebras, and was still unsolved in 1996. See [45] for details.
  - **Knowledge based systems**. Mycin (1975) was an expert system for bacterial infections and to recommend antibiotics. Prospector (1977) was a system for mineral exploration.
  - **Question Answering systems**. Watson participated in a match of Jeopardy! in 2011 and won the human participants.

2 Looking backwards

In this section we give an outline of the history of artificial intelligence, and give some hints on previous results that had influence or are in the roots of more recent developments.

2.1 Roots

Artificial intelligence has its roots in different disciplines. We list a few topics.

Logic and the interest of formalizing reasoning has a long history starting with Aristotle and his syllogisms. Syllogisms define correct inferences. The idea of automating reasoning appears in Ramon Llull. Leibniz, who knew Llull's work, worked with logic intensely. His goal with respect to logic was to build an universal calculus, to use it for arbitrary cientific propositions.

> Quo facto, quando orientur controversiae, non magis disputatione opus erit inter duos philosophos, quam inter duos Computistas. Sufficiet enim calamos in manus sumere sedereque ad abacos, et sibi mutuo (accito si placet amico) dicere: Calculemus (Leibniz 1875 [16])
> Translation: If this is done, whenever controversies arise, there will be no more need for arguing among two philosophers than among two matematicians. For it will suffice to take pens into the hand and to sit down by the abacus, saying to each other (and if they wish also to a friend called for help): Let us calculate. (Leibniz 1875 [16], translation from Lenzen 2004 [17])

Formalization of decision making and probability dates back to initial work on probabilities by Cardano (S.XVI), and later by other authors as Fermat, Pascal and Bernouilli in S. XVII and Laplace and Bayes in S. XVIII. Decision theory under uncertanty was studied by Bernouilli among others, but it was in 1944 when von Neumann and Morgenstern gave it a modern shape with their book [39]. It was also von Neumann who established the field of game theory in an article [37] a few years before (in 1928).

Reasoning and logic link AI to different sciences and areas as mathematics and epistemology, and philosophy. Decision theory connect AI with statistics, probability theory and economics. Neural networks links AI with neuroscience. Others relate with psychology, linguistic, control theory, and optimization.

2.2 Origins

The term artificial intelligence (AI) was used for the first time by J. McCarthy in 1955. It was used in the proposal to organize in summer 1956 the first meeting of researchers interested in this topic. It was the "Dartmouth Summer Research Conference on Artificial Intelligence", a two-month event that gathered researchers interested in intelligence, neural networks, and theory of automata. This meeting was organized by J. McCarthy (who was at Dartmouth College), M. L. Minsky (Harvard University), N. Rochester (IBM) and C. E. Shannon (Bell Telephone Laboratories). The participants were T. More (Princeton), A. Samuel (IBM), R. Solomonoff and O. Selfridge (MIT), and A. Newell and H. Simon (Carnegie Tech, now Carnegie Mellon University).

This meeting was a consequence of the first research in the area. Some of the tools already developed were discussed there. This was the case of the program Logic Theorist for automated reasoning developed by Newell and Simon. Logic Theorist was implemented in IPL (a language supporting lists) and it implemented reasoning as search, and for this purpose Newell and Simon had to introduce heuristics (to avoid combinatorial explosion).

In the same period, A. Samuel was developing his program for playing checkers. It introduced the minimax algorithm, alpha-beta pruning[2] to reduce search in games and an evaluation function (heuristics). C. Shannon in 1949 had also considered minimax and an evaluation function in a paper [30] on Chess.

These results can be seen as the origin of symbolic artificial intelligence (AI working with symbols).

We should mention another result from McCulloh and Pitts. They published in 1943 [21] a model of artificial neural networks, using logic (following Russell and Whitehead notation at the Principia Mathematica). This work can be seen as the origin of the area of neural networks within artificial intelligence. Although the model is based on logic, this work can be seen as the origin of non-symbolic articial intelligence.

The proposal [20] of the Dartmouth meeting enumerates "some aspects of the artificial intelligence problem". It includes references among other aspects to machine learning ("Self-Improvement") and neuron nets. These terms and the ones we have seen in this section are still the ones we find in AI systems today.

2.3 Early developments and difficulties: shaping the field

During the first years of the development of the field of AI, there was a lot of progress with programs that solved a large variety of different problems. Some of these programs are Analogy, Student, GPS (General Problem Solver, by Newell and Simon). Nevertheless, this progress was stopped around 1970. Russell and Norvig [26] give three reasons for the crisis.

1. **Systems had very little or no knowledge on the application domain**. Programs were mainly using transfomation rules. For example, the conversation system ELIZA, a program developed by Weizenbaum in mid 1960 to simulate a psychotherapist, had only a few syntactic transformation rules. Let us consider, for example, the following rules (see [40] for details).

   ```
   (I am (* x)) => (How long have you been (* x)?)
   ((* x) you (* y) me) => (what makes you think I (* y) me?)
   ```

   These rules permitted to reply the following sentence given by the user

   ```
   I am very unhappy these days.
   ```

   by the following one

   ```
   How long have you been unhappy these days?
   ```

[2] J. McCarthy and others are also credited for this invention.

Input1	Input2	XOR (Output)
0	0	0
0	1	1
1	0	1
1	1	0

Table 1: Exclusive or function.

The following sentence

`It seems that you hate me.`

can be replied by

`What makes you think I hate you?`

The only knowledge used in ELIZA was the one codified in rules like the previous ones.

The need to incorporate knowledge from the application domain, and to organize it so that its access was efficient and made systems for real applications possible fostered the development of knowledge representation formalisms.

Knowledge representation formalisms developed include semantic networks (Quillian, early sixties) and frames (Minsky, 1975). At the same time, logic was also considered for knowledge representation.

2. **Most of the problems were NP**. Because of that, they could be solved only for small instances, but when the dimension increased, they were not solvable.

   Most of the methods for problem solving were based on general purpose search mechanisms that chained different basic steps. The computational cost of these mechanisms is too high. First algorithms for search were introduced starting in 1959. Uniform-cost search was introduced by Dijkstra in 1959. Heuristic search was introduced by Newell and Ernst in 1965. Hart, Nilsson and Raphael developed A* search algorithm in 1968.

3. **Some of the basic structures used to solve the problem had limitations**. For example, the perceptrons (a type of neural network developed by Minsky and Papert) were not able to learn the exclusive or (xor) function (see Table 1). This is because they are not allowed to classify correctly non-linearly separable problems.

2.4 Some others not so early developments

In the first years of AI, there was also work to add learning capabilities to games. In 1962, E. B. Hunt describes the algorithm CLS to build non incrementally decision trees for classification. It is presented as a model of human learning. E. B. Hunt, J. Marin and P. T. Stone do in 1966 experiments on induction of concepts based on this algorithm. In 1979, Quinlan presents ID3, a method to build this decision trees. In 1970, P. Winston develops his program to learn descriptions of complex objects. In 1977, T. Mitchell develops the version space learning. All these methods fall in the area of machine learning, from a symbolic perspective. In the same period, the rediscovery of backpropagation, an algorithm for learning in neural networks, put this area of neural networks again in the mainstream research agenda.

Applications in real environments required to deal with aspects as uncertainty and imprecision (also needed for games of chance). Methods based on probability theory, and on alternative formalisms as fuzzy sets (L. Zadeh, 1965) and Dempster-Shafer (original from A. Demspter in 1968 and contributions of G. Shafer in 1976) were used.

All these results were used in the first commercial applications (mainly expert systems) in the 80s.

2.5 Present

The history of AI does not stop in the 80s. Methods and techniques have evolved since them, nevertheless the main areas in which we divide the field have not changed since this initial period.

In addition, some of the concepts that we see today are in direct debt with some of the first approaches. Current research topics relate with some of the ideas and concepts we have already seen.

For example, deep learning originates in neural networks, and data mining and knowledge discovery in machine learning. Data mining was introduced to use large amounts of data, large sets which now can be considered small with the appearance of big data in scene.

Big data introduce new problems, that need new solutions. Decision making based on streaming data is an example of this. New machine learning approaches have been developed for this purpose.

3 Why artificial intelligence?

We consider four motivations of why we study artificial intelligence. The first three ones are discussed by Scaruffi [27].

- Pure scientific curiosity. This links with research on mathematics, logics and philosophy. We have seen before that Leibniz considered that a universal calculus could be used for solving arguments. Hilbert, one of the developers of mathematical logic, wrote a list of open mathematical problems. One of them was the *Entscheidungsproblem* (decision problem). The problem was: given a statement in first-order logic, can we build an algorithm that states if the statement can be proven? A. Turing (using the Turing machine) and A. Church (using the lambda calculus) proved independently in 1935-1936 that the answer is no. Turing machines and lambda calculus are equivalent from a computational point of view[3].

So, a related problem is if we can build an algorithm that has an intelligent behavior. According to [42] Turing in 1945 had the "conviction that the computable operations were sufficient to embrace all mental functions performed by the brain", and studied "how computers could be made to perform operations which did not appear to be mechanical (to use common parlance)".

- Business. Automation improves productivity, and reduces costs.
- Idealistic. Some expect that knowledge based systems can provide expert advise when the experts are not available.
- Solve computational problems. There are problems for which we cannot write an efficient and optimal algorithm, either because this algorithm does not exist (e.g. it has been proven that the solution is NP) or because we do not know how to build it. Some AI tools can be used and are effective in such situations. Other sciences and fields in computer science have fostered the evolution of the field posing new problems. Search algorithms including evolutionary computation are useful for some of these problems. In the case that the problem is better understood and an efficient *standard* algorithm can be found, there is no need to use AI tools.

4 Definitions and points of view

There isn't a single definition of artificial intelligence, but different competing ones. Some of them are reviewed below together with the hypothesis behind AI systems.

4.1 Building programs

Artificial intelligence is a field in computer engineering, and as such, it is devoted to build programs. H. A. Simon states:

> The moment of the truth is a running program.
> (Simon 1995 [31], p. 96)

From this perspective, AI can be seen as an engineering.

[3] Turing machines are the models of imperative programming and lambda calculus the model of functional programming languages.

	human performance	rationality (ideal performance)
thought processes (reasoning)	thinking humanly (Cognitive model)	thinking rationally (Logics)
behavior	acting humanly (Turing test)	acting rationally

Table 2: The four types of systems according to the two dimensions (following Russell and Norvig [26] figure).

4.2 Goals of the programs

Which are the ultimate goals of AI programs are not so clear. Russell and Norvig classify goals in their book [26] according to two dimensions. They are the following ones.

1. The goal is to achieve a certain behavior (we are interested in the result of the program) or to reproduce a certain way of reasoning (we are more interested in how the result is obtained than in the result itself). According to this question, behaviour and reasoning define one dimension.

2. Programs have to be correct, but there are different ways to measure correctness. One is to compare programs with people, and the other is to compare them with an intelligence ideal. This ideal is known as rationality. The second dimension is thus about ways of measuring correctness.

Taking into account these two dimensions we have four different ways of defining artificial intelligence. All these four alternatives have been considered in the field of AI (see also Figure 2).

- Acting humanly: Turing test.

  A similar idea appears in McCarthy's proposal for the Darmouth meeting.

  > For the present purpose the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving. (McCarthy et al. 1955 [20])

- Thinking humanly. Cognitive model. See [36].

- Thinking rationally. Logics.

- Acting rationally.

4.3 Science and engineering

We have underlined above that artificial intelligence, as a field of computer engineering, is about the construction of programs. We can enlarge the development of software products considering that AI is also about (or related to) the design and construction of robots. All this underlyies the engineering part of AI.

However, at the same time, AI can also be seen as a science. Besides of developing new systems, it also studies intelligent systems and obtains new knowledge on them. According to H. A. Simon [31], it is an *empirical science* because we experiment. For example, Simon [31] (p. 99) states:

> Often the most efficient way to predict and understand the behavior of a novel complex system is to construct the system and observe it. Because AI programs are also computational models, we can use the programs themselves as their own models, an advantage for the field of AI that is unique in science. In AI, the theory not only *models* but simultaneously *exhibits* the behavior of the phenomena under study.
>
> The "natural" sciences also depend, for their progress, on building artificial systems and studying their behavior, for this is the essence of the experimental method. The natural scientist constructs a system in which the operation of certain natural law is thought to be prominent; then observes the phenomena produced by the system, and especially how these phenomena change with changes in the system parameters. So Galileo rolls balls down inclined planes or over the edges of tables, and measures the time of the roll or the length of the flight as a function of the angle of the plane.
>
> To experiment is to use the artificial to study the natural. To design an AI system and observe how its behavior changes with changes in the design is to perform an experiment. Most of what we know about artificial intelligence has been learned by carrying out experiments of this kind, thereby making AI a thoroughly experimental science. (Simon 1995 [31], p. 99)

As a summary, we can consider another quotation from Simon [31]:

> While the scientist is interested specifically in creating new knowledge, the engineer is interested also in creating systems that achieve desired goals. Apart from this difference in motives, there is no need to distinguish between computer scientists and computer engineers, or AI scientists and engineers. We can stop debating whether AI is science or engineering; it is both. (Simon 1995 [31])

This duality explains why AI literature ranges from applied research to theoretical results. The latter are needed to build systems as well as to explain the behavior of intelligent systems.

4.4 The physical symbol system hypothesis

Most of AI systems are developed under the so-called physical symbol system hypothesis, which was formulated by A. Newell and H. A. Simon [22].

> **The physical symbol system hypothesis.** A physical symbol system has the necessary and sufficient means for general intelligent action. (Newell and Simon 1976 [22])

Due to the fact that the resources are scarce, Newell and Simon considered that we need heuristic search to solve the problems. This hipothesis is the heuristic search hypothesis.

> **Heuristic Search Hypothesis.** The solutions to problems are represented as symbol structures. A physical symbol system exercises its intelligence in problem solving by search – that is, by generating and progressively modifying symbol structures until it produces a solution structure. (Newell and Simon 1976 [22])

The opinion of Newell and Simon is discussed by several authors within artificial intelligence. Some built models based on other hipothesis. We can underline three different alternatives based on three alternative models of intelligence. We discuss them below.

4.5 Sistems based on biological models

Neural networks are an alternative to physical symbols. They were designed according to biological models. A neural network is defined in terms of processing units, the neurons, and weighted connections between them. This definition does not include symbols. In addition, the knowledge is distributed in the network, and we cannot identify individual concepts in specific regions of the network. The same regions are used to represent several concepts.

Neural networks replace the focus on symbolic representation and on inference to emphasis by a focus on learning and adaptation. It is said that neural networks are mechanisms that adapt to the world (learning the weights in the connections) and there is no need to represent the world by means of symbols. Because of this, neural networks can be seen as a counterexample of the necessity of the hypothesis of physical symbols.

> **The physical symbol system hypothesis.** A physical symbol system has the necessary and sufficient means for general intelligent action. (Newell and Simon 1976 [22])

Although most authors consider that neural networks contradict the hypothesis of Newell and Simon, the latter has argued that we should consider the conditions for symbols in a broad sense so that it encompasses neural networks and R. Brooks robots (see text in Figure 1). In this way, the hypothesis is still valid. Under this premise neural networks can be seen as a knowledge representation formalism.

My work fits within the framework described above in terms of situatedness, embodiment, intelligence and emergence. In particular I have advocated situatedness, embodiment, and highly reactive architectures with no reasoning systems, no manipulable representations, no symbols, and totally decentralized computation. This different model of computation has lead to radically different models of thought.

Figure 1: Text from R. Brooks article, Intelligence without Reason [2] (Section 6).

> There is some dispute today about the Physical Symbol System Hypothesis, hinging on the definition of the term "symbol". If we define "symbol" narrowly, so that the basic components in connectionist systems or robots of the sort advocated by Brooks are not regarded as symbols, then the hypothesis is clearly wrong, for systems of these sorts exhibit intelligence. If we define symbols (as I have, above) as patterns that denote, then connectionist systems and Brooks robots qualify as physical symbol systems. (Simon 1995 [31], p.104-105).

4.6 Emergence

Another alternative to Newell and Simon's hypothesis comes from emergent intelligence. Using R. Brooks words [2] "intelligence emerges from the interaction of the components of the system", or, in general, it emerges from the activities and interactions of a set of independent agents. These activities can be rather simple although they can be specialized. Agents and multiagents technology shape distributed computing, and provide models of cooperation. We use individual agents to build complex systems using their cooperation skills. These ideas also appear in cognitive science where the mind is presumed to be organized as a set of specialized functional units.

So, in this case, we are contradicting the need of a single physical system.

> **The physical symbol system hypothesis.** A physical symbol system has the necessary and sufficient means for general intelligent action. (Newell and Simon [22])

Emergence is applied at different levels. We may have it at the individual/agent level, where the behavior of the agent is achieved by means of the interaction of independent modules, and at the population level. We have intelligence at the population level in the case of Brooks' robots, and in social intelligence in multiagent systems. Solutions are not built in a centralized way, it is the set of agents that define the behavior of the whole system.

4.7 Situatedness and embodiment

Situated action theories study artificial intelligence from a different perspective. Intelligence is not about the construction and evaluation of models about the world. It is a less structured process that focus on acting on the world, and on how we develop in it.

This approach gives more importance to actions than to explanations of the actions. So, behavior is more importance than reasoning. This consideration led Brooks to consider "Intelligence without representation" [3], which relates these systems to neural networks.

Reactive systems do not have internal state and they respond to stimuli of the environment. Then, importance is given to how the environment is sensed, and how stimuli are processed quickly to sense its changes. This led to give much importance to the body and the senses that situate the system into the world, instead of giving importance to higher level processes that reason on the environment.

In this theory, the implementation of the system is relevant. In order to act on the world, we need a physical body that permits the system to be situated in the world. Intelligence is the result of the activity of sensing everydays world. This is known as embodied intelligence.

Not all people that disagree on the physical symbols hypothesis agree on the need of embodiment. For example, H. M. Collins states that what is needed is social embedding. In this case, a physical body is not needed, instead we need interaction with social agents from the environment. Intelligence requires socialization. Collins states that the conceptual world is already situated.

> The odd thing about all of this is that Dreyfus, as a follower of Wittgenstein, ought not to think of the conceptual world as much different from the perceptual world. The conceptual world too is "situated". The later philosophy of Wittgenstein shows that what we take as logically and scientifically compelling is what we do not know how to doubt. What we take as certain is what follows for us, as a matter of course, in the way we live in the world. In the last resort there is no more compelling proof. Even logical syllogisms cannot be proved if we are unwilling just to "see" and act as though they follow in the situations in which we find ourselves. Thinking and acting in the world are but two sides of the same coin. (Collins 1996 [6], p. 101)

And also:

> In sum, the shape of the bodies of the members of a social collectivity and the situations in which they find themselves give rise to their form of life. Collectivities whose members have different bodies and encounter different situations develop different forms of life. But given the capacity for linguistic socialisation, an individual can come to share a form of life without having a body or the experience of

> physical situations which correspond to that form of life. What we dont know is how to make something with the capacity to be socialized in this way. (Collins 1996 [6], p. 105)

5 AI at present

There are a few topics that have appeared several times (as e.g. search, knowledge representation and machine learning) in the previous sections. They are the major areas of the field. All these lines are still progressing today. In addition to them, there is also the area of distributed artificial intelligence. On the one hand, this is about the exploitation of parallelism and distributed systems for solving problems. On the other, it is about the solution of problems that appear in distributed environments. Cooperative methods to be used by e.g. multiagent systems fall in this area.

As a summary, we can distinguish within AI the following major areas.

- Problem solving and search
- Knowledge representation and knowledge-based systems
- Machine learning
- Distributed artificial intelligence

In addition to them, there are a few *neighboring* areas that frequently are also considered as part of artificial intelligence. They are related to sensing and understanding the world, and acting on it. They include the following.

- Natural language
- Computer vision
- Robotics
- Speech recognition

5.1 Characteristics of AI systems

At present the number of AI systems is large, and there are large differences between systems. Nevertheless, there are a few characteristics that can be considered as representatives of them. We summarize them below.

- Symbolic information. The information they use and process is symbolic. They need to reason on abstract concepts, on facts, and they have to make conclusions from them.
- Domain knowledge. They need to know about their environment in order to make decisions, or find solutions. They need to know what is around, how they can act on the environment, and how the environment changes.

- Incomplete and uncertain data. Systems gather and process data that are not always *optimal* for their use. It is usual that the information is incomplete, there is uncertainty (randomness, imprecision, fuzziness) and that the pieces of information are in contradiction. Nevertheless, this information has to be used combining it with the knowledge in the system.

- Heuristic methods. Solutions are implemented using heuristic search. That is, a kind of guideline of how the solution has to be found, but that does not always ensure that we find the best solution, or any solution at all.

- Adaptive. Systems need to change their behavior when the environment changes. They need to learn from experience.

6 Elements for discussion

There are a few aspects on the foundations of AI that have led to discussion. As we have seen in Section 4, there are different points of view on what is and what should be AI. We present here some additional elements for discussion.

6.1 Understanding AI

6.1.1 Brute-force intelligence

Current AI systems usually exploit computational power as much as possible. Most of AI systems rely on the fact that computers permit us to evaluate millions of alternatives very quickly. Research tends to be on faster algorithms, on algorithms that better optimize results. Implementations use faster machines and exploit parallelism. This is known as brute-force AI.

- Search and games. Games are implemented using search algorithms, which mainly consists on the evaluation of different positions. For example, Hofstadter in an interview points out that Deep Blue evaluated 330 million positions a second while Kasparov few dozen [33]. With respect to playing Go, search algorithms were initially quite ineffective because in comparison with chess the number of positions to check was very large. To advance in the problem, research was directed to the combination of search with knowledge. Nevertheless, real progress was only achieved with a new search algorithm that minimizes knowledge. Other applications of search algorithms, as for solving constraint satisfaction problems (CSP), also rely on the huge computational power of computers. For example, [5] describes the system that schedules Hong Kong's Rail Network. It uses CSP. It is relevant to notice that these systems are much related to optimization and operations research.

- Machine translation (MT). This is another field in which current approaches are based on computational power and cheap memory. Statistical machine translation is the area studying this approach, and for example is the approach used by Google Translator. It is the approach mainly used today. See e.g. the following discussion.

> Statistical machine translation (SMT) is an approach to MT that is characterized by the use of machine learning methods. In less than two decades, SMT has come to dominate academic MT research, and has gained a share of the commercial MT market. (Lopez 2008 [18])

- Answering systems. Watson, the answering system that won Jeopardy! is another example of this approach. The system had milions of pages of information including "databases, taxonomies, and ontologies, such as dbPedia, WordNet, and the Yago8 ontology" [10].

- Machine learning. Some of the initial approaches to machine learning focused on knowledge and understanding. Current research trends focus on optimization (e.g., how to select parameters so that a model approximates some given data), and current research in machine learning is very much near to statistics.

As we have seen, there are a few areas of AI that converge with areas of statistics and operations research, as the knowledge used is minimal.

This trend on brute-force causes some controversy within AI. In short, the question is whether this is still artificial intelligence. J. McCarthy (see e.g. the quotation in Section 6.2.5 from [19]), Hofstadter (see e.g. [33]) and Scaruffi [27] are some of the people that have criticized this brute-force approach.

One of the consequences (and inconveniences) of this approach is that systems are usually useful on a very narrow application ninche. Another inconvenience (often related to the first) is that they are brittle. For example, Watson lost with the following question and answer.

- Question Its largest airport is named for a World War II hero; its second largest, for a World War II battle in the category of U.S. Cities,

- Watson said What is Toronto?

This was wrong, because Toronto is in Canada.

This error in the answer of Watson is discussed in [13].

If we want to classify the brute-force approach using the two dimensions above, we have that these systems are developed with the goal of *acting rationally*, that is, there is a function to be optimized and the AI program looks for a solution.

Note that we can reconsider Figure 2 to change the perspective given in [26] and underlining the need of thinking rationally. Figure 3 shows this change of perspective.

	human performance	rationality (ideal performance)
behavior	acting humanly (Turing test)	acting rationally
thought processes (reasoning)	thinking humanly (Cognitive model)	thinking rationally (Logics)

Table 3: The four types of systems according to the two dimensions but changing perspective.

6.1.2 Moravec's paradox

The trend toward brute-force is related to or complemented by the fact that "it is comparatively easy to make computers exhibit adult-level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility" (p. 28 [4]). This fact is known as Moravec's paradox.

These adult-level performance on intelligence tests and playing checkers are obtained by programs developed using the brute-force approach discussed in Section 6.1.1.

6.1.3 AI and understanding

- Chinese room (J. Searle)
- Strong AI vs. Weak AI.
  - Strong AI. AI systems can think. Artificial general intelligence. Systems can do any task humans can do. Searle
  - Weak AI. AI systems can only act as if they think.
- AI as Advanced Informatics (Searle)
- Machine translation. Does ELIZA know more with a few predefined syntactic tranformations than Watson or a machine learning translation system? Understands more?

6.2 AI and the society

6.2.1 Ethics and artificial intelligence

6.2.2 Privacy, pervarsive surveillance, and big data

Privacy is a fundamental right recognized in the Universal Declaratioon of Human Rights. Some countries have specific legislation for ensuring privacy. At the European level, in 2018 the new General Data Protection Regulation will enter into application. It replaces the data protection directive from 1995. Technology for ensuring some levels of privacy have been developed in the last 50 years.

Technologies for ensuring privacy were first considered within the statistical community, linked to national statistical offices and the need to ensure confidentiality to citizens in relation to census and questionnaires. Statistical disclosure control (SDC) is the area of statistics that studies methods for privacy of statistical databases.

Later, computer science studied also this topic in relation to communication and databases. The communities of privacy enhancing technologies (PETs) and privacy-preserving data mining (PPDM) were formed to study these topics within computer science. Data privacy [7, 15, 35] studies how to protect data, avoid disclosure while ensuring security and data utility, and quantify disclosure risk.

Big data poses several difficulties to data privacy. Some of this information is used for personalization and product customization, but also for profiling, advertisement, and stored for future (unknown) use. The amount of sensitive information available on people's economical translations, communications, and activities (e.g., trajectories tracked by telecom companies) increases every day. Besides of explicit sensitive information, another treat to privacy is that data mining algorithms permit to compute models from data that permit to infer sensitive information from other types of data that seem not so sensible. In addition, new services and products are constantly developed that still increase the amount of information stored.

A danger related to privacy is discrimination due to the application of machine learning algorithms without care.

The dangers of pervasive surveillance and lack of privacy are discussed in e.g. [25]. With respect to the first, it is stated that surveillance becomes political and is then a mechanism to protect the statu quo and maintain the political order.

6.2.3 Social responsability

The issue of social responsability is discussed in [25]. The paper focuses on cryptographers but most of the discussion is general for engineers. The responsability includes problem selection as "problem selection is the most obvious aspect in determining our community's impact" (p. 30, [25]). This discussion applies also to artificial intelligence.

6.2.4 Techno-optimists and techno-pessimists

There is an open debate on how current artificial intelligence technology (and computer engineering technology, in general) will affect society. There are the techno-optimists that think that the new technologies will have an extremely positive impact, causing an economic growth comparable or still larger than the one caused by the introduction of electricity. On the other hand, there are the techno-pessimists that are absolutely skeptical on the positive impact of these technologies. They consider that the economic growth that AI is causing will be negligible and that we are facing an era with no major economic growth.

R. J. Gordon in his long book [12] studies with detail production from 1870 to 1970 and argues that the production increase that started with the generalization of cheap energy (electricity) is now in its end. They argue that that technological revolution is not causing and will not cause a significant production increase. This book is centered to the economic growth in the United States, and while in other countries the economic growth may be posterior to 1970. According to Gordon, the cause is, anyway, cheap energy and not the digital society. The book also argues that the changes due to the electricity were more important than the ones caused by the digital computer[4].

Among the techno-optimists we find E. Brynjolfsson and A. McAfee. Their position is described in the book [4] titled "The second machine age". The book starts with a description of current technology where it discusses some of the limits of technology, including the description of Moravec's paradox cited in Section 6.1.2. Then, it discusses Gordon's [12] and Dutton's [8] ("The great stagnation") book.

Brynjolfsson and McAfee present "innovation-as-building-block" and state that "the number of potentially valuable building blocks is exploding around the world, and the possibilities are multiplying as never before" (p.81, [4]). Internet, the access to scientific literature, advanced software, and powerful computers will permit to people (and computers) all around the planet to evaluate these possibilities. Then, "with more eyeballs, more powerful combinations will be found" (p.83, [4]). This will be the basis of the new growth that is now just starting.

Nevertheless, these two techno-optimists also express they concern about a few complex problems that the United States face today. Some of them are, in fact, not restricted to U.S. They show that the mean income adjusted for inflation started in 1979 to have small increases and later, from 1999, decreases; that mean and median income started to have different behaviours (p. 132); and that while before all salaries had similar behaviours (increase or decrease similarly) this is not longer the case. They show that high school dropout have now a salary lower (adjusted by inflation) than the one in 1963 (p. 136) and that only "college graduates" and "graduate school" have now better salaries (inflation-adjusted) than the ones of 1973. The book discusses about work polarization "a collapse in demand for middle-income jobs, while nonroutine cognitive jobs (such as financial analysis) and nonroutine manual jobs (like hair-dressing) have held up relatively well" (p. 139); and about super stars with super salaries caused by the "winner-take-all-society". The global market (a side-effect of internet) is said to be the cause of this new model, where salaries do not follow a normal distribution but a power-law distribution (p.160).

The book also discusses if the fact that salaries have decreased is really a

[4]Compare this Scarusffi's text [27] p.22: "We also have to realize that what looks astonishing to us today might not look so impressive to future generations. ... When the car was invented, instead, it must have been an incredible sight: an object that moves with no animal pulling it and no peasants pushing it. (...) Today, lick back then, the temptation is to call it 'magic' (or 'Artificial Intelligence'). In reality, it is just an application of well-known physical laws."

bad thing. The "strong bounty argument" states that this is not a big problem. Decrease of salary is compensated by the fact that the quality of services and products is much better. Brynjolfsson and McAfee, however, argue that this is not the case and in general the current situation for people with less salary is worse than in 1999. The book follow with a few short and long term political proposals.

From the AI perspective, it is specially relevant a discussion on the race humankind against artificial intelligence. The authors state that this race is not over. Precisely the book gives the example of chess, and the freestyle chess where players can play with the help of computers. The authors state that in this case, human players win computer programs. Not so good human players with not so good computer programs are able to win the best computer programs. The authors conclude this discussion with: "We've never seen a truly creative machine" (p. 191 [4]).

As we have shown, [4] discusses about the quality of life today in comparison with the one in the late 90s for the working class with lower salaries. Topol, in his book [34], expresses his techno-optimism in medicine. He compares the smartphone to Gutenberg's press, for medicine and patienr treatment.

[27] discusses a related issue. He argues that some of the products today are not of better quality than the ones prior to the digital revolution, and that computer engineering has some harmful side effects. Consider the following quotations:

- "People tend to believe machines more than they believe humans, and, surprisingly, seem to trust machine-mediated opinions better than first-hand opinions from an expert" (p. 121, [27]);
- "One wonders whether society is aiming for the technology that minimizes our responsibility instead of aiming for the technology that maximizes our effectiveness" (p. 121, [27]);
- "Computer science is becoming the discipline of turning your life into somebody else's business" (p. 126, [27]).

6.2.5 Theory and practice in research

> If you know where you'r going, you'r not gona find anything really interesting. (Michael Levitt, Nobel Prize in Chemistry 2013)

There has always been discussion on whether research has to be applied, or if basic research is useful, rellevant, and has to be supported. In this aspect, artificial intelligence is not different from other sciences and the debate has also arisen here.

John McCarthy, when he was president of the Association for the Advancement of Artificial Intelligence (AAAI) made an article [19] advocating for more emphasis on basic research. The title of the article already underlines his point " AI Needs more emphasis on basic research". According to McCarthy, basic research includes research on long terms goals of AI. In this paper there is also a critical comment against chess research trends (this paper was written in 1983, well before Deep Blue won Kasparov):

> Unfortunately, chess was discouraged as a serious problem domain, and most chess programming is carried on at the level of sport rather than science. In particular, there is little publication about the intellectual mechanisms involved, and the race often involves merely faster hardware. (McCarthy 1983 [19], p. 5)

Support for basic research is not always supported by funding agencies. Some researchers are also against this support, or advocate for a priorization of applied research. See e.g. the words of John R. Rice in [24] were he discusses the future of computing research.

> A computing researcher who cannot connect to strategic research in a direct way probably does not deserve research funding. If one does not have enough creativity to make this connection, then one probably does not have enough creativity to do significant research either. (Rice 1995 [24])

This is an old discussion, Plutarch states in his Parallel Lives that Archimedes believed that were his results on pure mathematics the superior ones, and not his inventions (see e.g. [44]). Much later (1808), C. F. Gauss (1777-1855) mentioned Archimedes and a poem by Friedrich Schiller (1759-1805) titled "Archimedes und der Schüler" (Figure 2) for similar reasons.

> They searched the truth for its own sake and found in the very success of their efforts their reward and their happiness. I cannot avoid at this point reminding you of Archimedes, who was admired by his contemporaries mainly just because of his artful machines, because of the apparently magical workings they had, but who placed so little value on all this, compared with his magnificent discoveries in the field of pure mathematics which in themselves usually had no visible benefits in the common sense of the term, at least then that he did not write down for us anything about the former, while he developed the latter with affection in his immortal works. You must all know the beautiful poem by Schiller. Let us consider also the sublime astronomy from this beautiful standpoint above all. (Gauss 1808 [11] translation from Ferreiros 2006 [9])

In contrast, J. B. J. Fourier (1768-1830) considered that mathematics was valuable because of its applications.

John A. von Neumann in his address in 1954 [38], which starts reminding Friedrich Schiller's poem, concludes as follows.

> This is true for all science. Successes were largely due to forgetting completely about what one ultimately wanted, or whether one

Archimedes und der Schüler by Friedrich Schiller

Zu Archimedes kam ein wißbegieriger Jüngling.
,,Weihe mich," sprach er zu ihm, ,,ein in die göttliche Kunst,
Die so herrliche Frucht dem Vaterlande getragen
Und die Mauern der Stadt vor der Sambuca beschützt¡'
,,Göttlich nennst du die Kunst? Sie ist's," versetzte der Weise;
,,Aber das war sie, mein Sohn, eh sie dem Staat noch gedient.
Willst du nur Früchte von ihr, die kann auch die sterbliche zeugen;
Wer um die Göttin freit, suche in ihr nicht das Weib."

Translation: **Archimedes and the Student** by Friedrich Schiller

To Archimedes came a youth desirous of knowledge.
"Tutor me," spoke he to him, "in the most godly of arts,
Which such glorious fruit to the land of our father hath yielded
And the walls of the town from the Sambuca preserved!"
"Godly nams't thou the art? She ist," responded the wise one;
"But she was that, my dear son, e're she the state served.
Woulds't thou but the fruits from her, these too can the mortal engender;
Who doth woo the Goddess, seek not the woman in her."

Figure 2: Friedrich Schiller poem *Archimedes und der Schüler* from [41]. Translation from [47] by William F. Wertz.

> wanted anything ultimately; in refusing to investigate things which profit, and in relying solely on guidance by criteria of intellectual elegance; it was by following this rule that one actually got ahead in the long run, much better than any strictly utilitarian course would have permitted. (von Neumann 1954 [38], p. 489)

See the references above, Hardy's book [14] "A Mathematician's Apology, and [32] for more details and discussion.

References

[1] Bostrom, N., Yudkowsky, E. (2011) The ethics of artificial intelligence, in W. Ramsey, K. Frankish (eds.) Handbook of Artificial Intelligence, Cambridge University Press, 2011.

[2] Brooks, R. A. (1991) Intelligence without reason, MIT AI Memo No. 1293.

[3] Brooks, R. A. (1991) Intelligence without representation, Artificial Intelligence 47 139-159.

[4] Brynjolfsson, E., McAfee, A. (2016) The second machine age, W. W. Norton & Company, Inc.. Paperback edition 2016.

[5] Chun, A. H. W., Suen, T. Y. T. (2014) Engineering works scheduling for Hong Kong's rail network, Proc. 26th Annual Conf. Innovative Applicatins of AI.

[6] Collins, H. W. (1996) Embedded or embodied? a review of Hubert Dreyfus' What computers still can't do, Artificial Intelligence 80 99-117.

[7] Duncan, G. T., Elliot, M., Salazar, J. J. (2011) Statistical confidentiality, Springer.

[8] Cowen, T. (2011) The great stagnation, Dutton.

[9] Ferreiros, J. (2006) The rise of pure mathematics as arithmetic with Gauss, in C. Goldstein, N. Schappacher, J. Schwermer (eds) The Shaping of Arithmetic: Number theory after Carl Friedrich Gausss Disquisitiones Arithmeticae, Springer, Berlin.

[10] Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A. A., Lally, A., Murdock, J. W., Nyberg, E., Prager, J., Schlaefer, N., Welty, C. (2010) Building Watson: An Overview of the DeepQA Project, AI Magazine.

[11] Gauss, Carl Friedrich (1808) Astronomische Antrittsvorlesung in C. F. Gauss (1929) Werke, vol. XII, Varia. Atlas des Erdmagnetismus, ed. Königliche Gesellschaft der Wissenschaften zu Göttingen. Berlin: Julius Springer (Repr. Hildesheim, New York: Olms, 1981), pp. 177-198.

[12] Gordon, R. J. (2016) The rise and fall of American growth, Princeton University Press.

[13] Hamm, S. (2011) Watson on Jeopardy! Day Two: The Confusion over an Airport Clue, post February, 15th 2011 19:30
`http://asmarterplanet.com/blog/2011/02/watson-on-jeopardy-day-two-the-confusion-over-an`

[14] Hardy, G. (1940) A mathematician's apology, Cambridge University Press.

[15] Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E. S., Spicer, K., de Wolf, P.-P. (2012) Statistical Disclosure Control, Wiley.

[16] Leibniz, G. W. (1875) De scientia universalis seu calculo philosophico, in C. I. Gerhardt (ed.) Die Philosophischen Schriften von Gottfried Wilhelm Leibniz, vol. VII, Weidmann.

[17] Lenzen, W. (2004) Leibniz's logic, in D. M. Gabbay, J. Woods (eds.) Handbook of the History of Logic vol. 3, 1-83.

[18] Lopez, A. (2008) Statistical machine translation, ACM Computing Surveys 40:3.

[19] McCarthy, J. (1983) AI Needs more emphasis on basic research, AI Magazine 4:4 p.5

[20] McCarthy, J., Minksy, M. L., Rochester, N., Shannon, C. E. (1955) A proposal for the Dartmouth summer research project on artificial intelligence.

[21] McCulloch, W. S., Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity, B. of Mathematical Biophysics 5 115-133.

[22] Newell, A., Simon, H. A. (1976) Computer Science as Empirical Inquiry: Symbols and Search, Communications ACM 19:3 113-126.

[23] Pritchard, D. (2014) What is this thing called knowledge, Routledge.

[24] Rice, J. R. (1995) Computational science and the future of computing research, Purdue University, CSD-TR-95-045.

[25] Rogaway, P. (2016) The Moral Character of Cryptographic Work. Essay written to accompany an invited talk (the 2015 IACR Distinguished Lecture) given at Asiacrypt 2015.
`http://web.cs.ucdavis.edu/~rogaway/papers/moral-fn.pdf`

[26] Russell, R., Norvig, P. (1995) Artificial Intelligence: A Modern Approach, Prentice-Hall. 1st. Edition 1995. 2nd Edition 2002. 3rd Edition 2010.
`http://aima.cs.berkeley.edu/`

[27] Scaruffi, P. (2013) Demystifying machine intelligence, Omniware.

[28] Schaeffer, J. (2008) One jump ahead: challenging human supremacy in checkers, Springer (2nd edition) (1st edition 1997)

[29] Schaeffer, J., Burch, N., Bjornsson, Y., Kishimoto, A., Muller, M., Lake, R., Lu, P., Sutphen, S. (2007) Checkers is Solved, Science 317 1518-1522.

[30] Shannon, C. E. (1950) Programming a Computer for Playing Chess, Philosophical Magazine, Ser.7, 41:314, March. Presented at the National IRE Convention, March 9, 1949, New York, U.S.A.

[31] Simon, H. A. (1995) Artificial intelligence: an empirical science, Artificial intelligence 77 95-127.

[32] Smith, J. T. (2014) Hilbert's Radio Address - Hilbert's Sources - Their Own Words, Convergence, June.
http://www.maa.org/press/periodicals/convergence/hilberts-radio-address-hilberts-sources-their-own-words

[33] Sommers, J. (2013) The Man Who Would Teach Machines to Think, The Atlantic.
`http://www.theatlantic.com/magazine/archive/2013/11/the-man-who-would-teach-machines-to`

[34] Topol, E. (2015) The patient will see you, Basic Books.

[35] Torra, V. (2017) Data privacy, in preparation.

[36] Vernon, D. (2014) Artificial Cognitive Systems - A Primer, MIT Press.

[37] von Neumann, J. (1928) Zur Theorie der Gesellschaftsspiele, Mathematische Annalen 100:1 295-320

[38] von Neumann, J. A. (1954) The role of mathematics in the sciences and in society, address before Assoc. of Princeton Graduate Alumni, June 16.

[39] von Neumann, J., Morgenstern, O. (1944) The theory of games and economic behavior, Princeton University Press.

[40] Weizenbaum, J. (1966) ELIZA – A computer program for the study of natural language communication between man and machine

URL and Web Pages

[41] `http://gutenberg.spiegel.de/buch/friedrich-schiller-gedichte-3352/12`

[42] `http://plato.stanford.edu/entries/turing/`

[43] `http://webdocs.cs.ualberta.ca/~chinook/`

[44] `http://www-history.mcs.st-and.ac.uk/Biographies/Archimedes.html`

[45] http://www.cs.unm.edu/ mccune/papers/robbins/

[46] `http://www.research.ibm.com/deepblue`

[47] `http://www.schillerinstitute.org/educ/pedagogy/list.html#schiller`