Artificial Intelligence: A Modern Approach

By Stuart J. Russell et all.

First Edition

CONTENTS

CHAPTER 1

INTRODUCTION

We call ourselves *Homo sapiens* – man the wise – because our **intelligence** is so important to us. For thousands of years, we have tried to understand *how we think*; that is, how a mere handful of matter can perceive, understand, predict, and manipulate a world far larger and more complicated than itself. The field of **artificial intelligence**, or AI, goes further still: it attempts not just to understand but also to *build* intelligent entities.

1 What Is AI?

In Figure 1.1 we see eight definitions of AI, laid out along two dimensions. The definitions on the left measure success in terms of fidelity to *human* performance, whereas the ones on the right measure against an *ideal* performance measure, called **rationality**.

Thinking Humanly	**Thinking Rationally**
"The exciting new effort to make computers think ... *machines with minds*, in the full and literal sense." (Haugeland, 1985) "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem-solving, learning ..." (Bellman, 1978)	"The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985)
Acting Humanly	**Acting Rationally**
"The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990) "The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)	"Computational Intelligence is the study of the design of intelligent agents." (Poole *et al.*, 1998) "AI ... is concerned with intelligent behavior in artifacts." (Nilsson, 1998)

Figure 1.1: Some definitions of artificial intelligence, organized into four categories.

1.1 Acting humanly: The Turing Test approach

The **Turing Test**, proposed by Alan Turing (1950), was designed to provide a satisfactory operational definition of intelligence. For now, we note that programming a computer to pass a rigorously applied test provides plenty to work on. The computer would need to possess the following capabilities:

- **natural language processing** to enable it to communicate successfully in English.
- **knowledge representation** to store what it knows or hears;
- **automated reasoning** to use the stored information to answer questions and to draw new conclusions.

- **machine learning** to adapt to new circumstances and to detectand extrapolate patterns.

Turing's test deliberately avoided direct physical interaction between the interrogator and the computer, because *physical* simulation of a person is unnecessary for intelligence. However, the so-called **total Turing Test** includes a video signal so that the interrogator can test the subject's perceptual abilities, as well as the opportunity for the interrogator to pass physical objects "through the hatch." To pass the total Turing Test, the computer will need

- **computer vision** to perceive objects, and
- **robotics** to manipulate objects and move about.

1.2 Thinking humanly: The cognitive modeling approach

The interdisciplinary field of **cognitive science** brings together computer models from AI and experimental techniques from psychology to construct precise and testable theories of the human mind.

1.3 Thinking rationally: The "laws of thought" approach

The Greek philosopher Aristotle was one of the first to attempt to codify right thinking, that is, irrefutable reasoning processes. His **syllogisms** provided patterns for argument structures that always yielded correct conclusions when given correct premises. These laws of thought were supposed to govern the operation of the mind; their study initiated the field called **logic**.

By 1965, programs existed that could, in principle, solve *any* solvable problem described in logical notation. (Although if no solution exists, the program might loop forever.) The so-called **logicist** tradition within artificial intelligence hopes to build on such programs to create intelligent systems.

1.4 Acting rationally: The rational agent approach

An **agent** is just something that acts (*agent* comes from the Latin *agere*, to do). Of course, all computer programs do something, but computer agents are expected to do more: operate autonomously, perceive their environment, persist over a prolonged time period, adapt to change, and create and pursue goals. A **rational agent** is one that acts so as to achieve the best outcome or, when there is uncertianty, the best expected outcome.

2 The Foundations of Artificial Intelligence

2.1 Philosophy

Descartes was a strong advocate of the power of reasoning in understanding the world, a philosophy now called **rationalism**, and one that counts Aristotle and Leibnitz as members. But Descartes was also a proponent of **dualism**. An alternative to dualism is **materialism**, which holds that the brain's operation according to the laws of physics *constitutes* the mind.

The **empiricism** movement, starting with Francis Bacon's (1561-1626) *Novum Organum*, is characterized by a dictum of John Locke (1632-1704):"Nothing is in the understanding, which was not first in the senses." David Hume's (1711-1776) *A Treatise of Human Nature* (Hume, 1739) proposed what is now known as the principle of **induction**: the general rules are acquired by exposure to repeated associations between their elements. Building on the work of Ludwig Wittgenstein (1889-1951) and Bertrand Russell (1872-1970), the famous Vienna Circle, led by Rudolf Carnap (1891-1970), developed the doctrine of **logical positivism**. This doctrine holds that all knowledge can be characterized by logical theories connected, ultimately, to **observation sentences** that correspond to sensory inputs; thus logical positivism combines rationalism and empiricism. The **confirmation theory** of Carnap and Carl Hempel (1905-1997) attempted to analyze to acquisition of knowledge from experience.

2.2 Mathematics

The first nontrivial **algorithm** is thought to be Euclid's algorithm for computing greatest common divisors. In 1930, Gödel showed that limits on deduction do exist. His **incompleteness theorem** showed that in any formal theory as strong as Peano arithmetic (the elementary theory of natural numbers), there are true statements that are undecidable in the sense that they have no proof within the theory.

The fundamental result can also be interpreted as showing that some functions on the integers cannot be represented by an algorithm – that is, they cannot be computed. This motivated Alan Turing (1912-1954) to try to characterize exactly which functions *are* **computable** – capable of being computed.

Although decidability and computability are important to an understanding of computation, the notion of **tractability** has had a even greater impact.

How can one recognize an intractable problem? The theory of **NP-completeness**, pioneered by Steven Cock (1971) and Richard Karp (1972), provides a method.

Besides logic and computation, the third great contribution of mathematics to AI is the theory of **probability**.

2.3 Economics

- How should we make decisions so as to maximize payoff?
- How should we do this when others may not go along?
- How should we do this when the payoff may be far in the future.

The mathematical treatment of "preferred outcomes" or **utility** was first formalized by Léon Walras (pronounced "Valrasse")(1834-1910) and was improved by Frank Ramsey (1931) and later by John von Neumann and Oskar Morgenstern in their book *The Theory of Games and Economic Behavior* (1944).

Decision theory, which combines probability theory with utility theory, provides a formal and complete framework for decisions (economic or otherwise) made under uncertainty – that is, in cases where probabilistic descriptions appropriately capture the decision maker's environment. For "small" economies, the situation is much more like a **game**: the actions of one player can significantly affect the utility of another (either positively or negatively). Von Neumann and Morgenstern's development of **game theory** (see also Luce and Raiffa, 1957) included the surprising result that, for some games, a rational agent should adopt policies that are (or least appear to be) randomized.

For the most part, economists did not address the third question listed above, namely, how to make rational decisions when payoffs from actions are not immediate but instead result from several actions take *in sequence*. This topic was pursued in the field of **operations research**, which emerged in World War II from efforts in Britain to optimize radar installations, and later found civilian applications in complex management decisions. The work of Richard Bellman (1957) formalized a class of sequential deicison problems called **Markov decision processes**.

The pioneering AI researcher Herbert Simon (1916-2001) won the Nobel Prize in economics in 1978 for his early work showing that models based on **satisficing** – making decisions that are "good enough," rather than laboriously calculating an optimal decision – gave a better description of actual human behavior (Simon, 1947).

2.4 Neuroscience

Neuroscience is the study of the nervous system, particularly the brain.

Paul Broca's (1824-1880) study of aphasia (speech deficit) in brain-damaged patients in 1861 demonstrated the existence of localized areas of the brain responsible for specific cognitive functions. By that time, it was known that the brain consisted of nerve cells, or **neurons**, but it was not until 1873 that Camillo Golgi (1843-1926) developed a staining technique allowing the observation of individual neurons in the brain.

Figure 1.2 shows that computers have a cycle time that is a million times faster than a brain. Futurists make much of these numbers, pointing to an approaching **singularity** at which computers reach a superhuman level of performance (Vinge, 1993; Kurzweil, 2005), but the raw comparisons are not especially informative.

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

Figure 1.2: A crude comparison of the raw computational resources available to the IBM BLUE GENE supercomputer, a typical personal computer of 2008, and the human brain. The brain's numbers are essentially fixed, whereas the supercomputer's numbers have been increasing by a factor of 10 every 5 years or so, allowing it to achieve rough parity with the brain. The personal computer lags behind on all metrics except cycle time.

2.5 Psychology

Wundt insisted on carefully controlled experiments in which his workers would perform a perceptual or associative task while introspecting on their thought processes. Biologists studying animal behavior, on the other hand, lacked introspective data and developed an objective methodology, as described by H.S. Jennings (1906) in his influential work *Behavior of the Lower Organisms*. Applying this viewpoint to humans, the **behaviorism** movement, led by John Watson (1878-1958), rejected *any* theory involving mental processes on the grounds that introspection could not provide reliable evidence.

Cognitive psychology, which views the brain as an information-processing device, can be traced back at least to the works of William James (1842-1910). After Craik's death in a bicycle accident in 1945, his work was continued by Donald Broadbent, whose book *Perception and Communication* (1958) was one of the first works to model psychological phenomena as information processing. Meanwhile, in the United States, the development of computer modeling led to the creation of the field of **cognitive science**.

2.6 Control theory and cybernetics

The central figure in the creation of what is now called **control theory** was Norbert Wiener (1894-1964). Ashby's *Design for a Brain* (1948, 1952) elaborated on his idea that intelligence could be created by the use of **homeostatic** devices containing appropriate feedback loops to achieve stable adaptive behavior.

Modern control theory, especially the branch known as stochastic optimal control, has as its goal the design of systems that maximize an **objective function** over time.

2.7 Linguistics

Chomsky pointed out that the behaviorist theory did not address the notion of creativity in language – it did not explain how a child could understand and make up sentences that he or she had never heard before. Chomsky's theory – based on syntactic models going back to the Indian linguist Panini (c. 350 B.C.) – could explain this, and unlike previous theories, it was formal enough that it could in principle be programmed.

Modern linguistics and AI, then, were "born" at about the same time, and grew up together, intersecting in a hybrid field called **computational linguistics** or **natural language processing**. Much of the early work in **knowledge representation** (the study of how to put knowledge into a form that a computer can reason with) was tied to language and informed by research in linguistics, which was connected in turn to decades of work on the philosophical analysis of language.

3 The History of Artificial Intelligence

3.1 The gestation of artificial intelligence (1943-1955)

Donald Hebb (1949) demonstrated a simple updating rule for modifying the connection strengths between neurons. His rule, now called **Hebbian learning**, remains an influential model to this day.