The Foundations of Artificial Intelligence

Philosophy

Can formal rules be used to draw valid conclusions?

How does the mental mind arise from a physical brain?

Where does knowledge come from?

How does knowledge lead to action?

Aristotle (384-322 B.C.) was the first to formulate a precise set of laws governing the rational part of the mind. He developed an informal system of syllogisms for proper reasoning, which in principle allowed one to generate conclusions mechanically, given initial premises.

Much later, Ramon Lull (d. 1315) had the idea that useful reasoning could actually be carried out by a mechanical artifact – “concept wheels“.

Thomas Hobbes (1588-1679) proposed that reasoning was like numerical computation that “we add and subtract in our silent thoughts.” The automation of computation itself was already well under way.

Around 1500, Leonardo da Vinci (1452-1519) designed but did not build a mechanical calculator; recent reconstructions have shown the design to be functional.

The first known calculating machine was constructed around 1623 by the German scientist Wilhelm Schickard (1592-1635).

The Pascaline, built in 1642 by Blaise Pascal (1623-1662), is more famous. Pascal wrote that “the arithmetical machine produces effects which appear nearer to thought than all the actions of animals.”

Gottfried Wilhelm Leibniz (1646-1716) built a mechanical device intended to carry out operations on concepts rather than numbers, but its scope was rather limited.

Now that we have the idea of a set of rules that can describe the formal, rational part of the mind, the next step is to consider the mind as a physical system.

Rene Descartes (1596-1650) gave the first clear discussion of the distinction between mind and matter and of the problems that arise.

One problem with a purely physical conception of the mind is that it seems to leave little room for free will: if the mind is governed entirely by physical laws, then it has no more free will than a rock “deciding” to fall toward the center of the earth.

Problems with Cartesian dualism

Alternative to dualism is materialism, which holds that the brain’s operation according to the laws of physics constitutes the mind. Free will is simply the way that the perception of available choices appears to the choice process.

Given a physical mind that manipulates knowledge, the next problem is to establish the source of knowledge. The empiricism movement, starting with Francis Bacon’s (1561-1626) Novum 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**: that 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.

The **confirmation theory** of Carnap and Carl Hempel (1905-1997) attempted to understand how knowledge can be acquired from experience.

Carnap’s book *The Logical Structure of the World* (1928) defined an explicit computational procedure for extracting knowledge from elementary experiences. It was probably the first **theory of mind as a computational process**.

The final element in the philosophical picture of the mind is the **connection between knowledge and action**. This question is vital to AI, because intelligence requires action as well as reasoning. Moreover, only by understanding how actions are justified can we understand how to build an agent whose actions are justifiable (or rational).

Aristotle argued that actions are justified by a logical connection between goals and knowledge of the action’s outcome. Aristotle’s algorithm was implemented 2300 years later by Newell and Simon in their GPS program. Goal-based analysis is useful, but does not say what do when several actions will achieve the goal, or when no action will achieve it completely.

**Mathematics**

**What are the formal rules to draw valid conclusions?**

**What can be computed?**

**How do we reason with uncertain information?**

Philosophers staked out most of the important ideas of AI, but the leap to a formal science required a level of **mathematical formalization** in three fundamental areas: logic, computation, and probability.

The idea of formal logic can be traced back to the philosophers of ancient Greece, but its mathematical development really began with the work of George Boole (1815-1864), who worked out the details of **propositional or Boolean logic**.

In 1879, Gottlob Frege (1848-1925) extended Boole’s logic to include objects and relations, creating the **first-order logic** that is used today as the most basic knowledge representation system.

**Alfred Tarski** (1902-1983) introduced a theory of reference that shows how to relate the objects in a logic to objects in the real world.

The next step was to determine the limits of what could be done with logic and computation.
The first nontrivial algorithm is thought to be Euclid’s algorithm for computing greatest common denominators.

The study of algorithms as objects in themselves goes back to al-Khowarazmi, a Persian mathematician of the 9th century, whose writings also introduced Arabic numerals and algebra to Europe.

Boole and others discussed algorithms for logical deduction and by the late 19th century, efforts were under way to formalize general mathematical reasoning as logical deduction.

In 1900, David Hilbert (1862-1943) presented a list of 23 problems that he correctly predicted would occupy mathematicians for the bulk of the century.

The final problem asks whether there is an algorithm for deciding the truth of any logical proposition involving the natural numbers – the famous “decision problem”. Essentially, Hilbert was asking whether there were fundamental limits to the power of effective proof procedures.

In 1930, Kurt Godel (1906-1978) showed that there exists an effective procedure to prove any true statement in the first-order logic of Frege and Russell, but that first-order logic could not capture the principle of mathematical induction needed to characterize the natural numbers.

In 1931, he showed that real limits do exist. His incompleteness theorem showed that in any language expressive enough to describe the properties of the natural numbers, there are true statements that are undecidable in the sense that their truth cannot be established by any algorithm.

This fundamental result can also be interpreted as showing that there are some functions on the integers that 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 capable of being computed. This notion is actually slightly problematic, because the notion of a computation or effective procedure really cannot be given a formal definition.

However, the Church-Turing thesis, which states that the Turing machine (Turing, 1936) is capable of computing any computable function, is generally accepted as providing a sufficient definition.

Turing also showed that there were some functions that no Turing machine can compute. For example, no machine can tell in general whether a given program will return an answer on a given input or run forever.

Although undecidability and noncomputability are important to an understanding of computation, the notion of intractability has had a much greater impact.

A problem is called intractable if the time required to solve instances of the problem grows exponentially with the size of the instances.

Besides logic and computation, the third great contribution of mathematics to AI is the theory of probability. The Italian Gerolamo Cardano (1501-1576) first framed the idea of probability, describing it in terms of the possible outcomes of gambling events.

Probability quickly became an invaluable part of all the quantitative sciences, helping to deal with uncertain measurements and incomplete theories.
Pierre Fermat (1601-1665), Blaise Pascal (1623-1662), James Bernoulli (1654-1705), Pierre Laplace (1749-1827), and others advanced the theory and introduced new statistical methods. Thomas Bayes (1702-1 761) proposed a rule for updating probabilities in the light of new evidence. Bayes’ rule and the resulting field called Bayesian analysis form the basis of most modern approaches to uncertain reasoning in AI systems.

Economics (1776-present)

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 science of economics got its start in 1776, when Scottish philosopher Adam Smith (1723-1790) published An Inquiry into the Nature and Causes of the Wealth of Nations. While the ancient Greeks and others had made contributions to economic thought, Smith was the first to treat it as a science, using the idea that economies can be thought of as consisting of individual agents maximizing their own economic well-being.

Most people think of economics as being about money, but economists will say that they are really studying how people make choices that lead to preferred outcomes or utility.

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.

This is suitable for “large” economies where each agent need pay no attention to the actions of other agents as individuals.

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 included the surprising result that, for some games, a rational agent should act in a random fashion, or at least in a way that appears random to the adversaries.

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 taken 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 decision problems called Markov decision processes.

Work in economics and operations research has contributed much to our notion of rational agents.

For many years AI research developed along entirely separate paths. One reason was the apparent complexity of making rational decisions.
Herbert Simon, the pioneering AI researcher, 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. The earlier gives a better description of actual human behavior.

Neuroscience (1861-present)

How do brains process information?

Neuroscience is the study of the nervous system, particularly the brain. The exact way in which the brain enables thought is one of the great mysteries of science.

It has been appreciated for thousands of years that the brain is somehow involved in thought, because of the evidence that strong blows to the head can lead to mental incapacitation.

Brain is recognized as the seat of consciousness. Before then, candidate locations included the heart, the spleen, and the pineal gland.

Paul Broca’s (1824-1880) study of aphasia (speech deficit) in brain-damaged patients in 1861 reinvigorated the field and persuaded the medical establishment of the existence of localized areas of the brain responsible for specific cognitive functions.

Despite these advances, we are still a long way from understanding how any of these cognitive processes actually work.

The truly amazing conclusion is that a collection of simple cells can lead to thought, action, and consciousness or, in other words, that brains cause minds (Searle, 1992).

The only real alternative theory is mysticism: that there is some mystical realm in which minds operate that is beyond physical science.

Brains and computers perform quite different tasks and have different properties.

Moore’s Law predicts that the CPU’s gate count will equal the brain’s neuron count around 2020. Moore’s Law says that the number of transistors per square inch doubles every 1 to 1.5 years. Human brain capacity doubles roughly every 2 to 4 million years.

Even though a computer is a million times faster in raw switching speed, the brain ends up being 100,000 times faster at what it does.

Psychology (1879-present)

How do humans and animals think and act?

The origins of scientific psychology are usually traced to the work of the German physicist Hermann von Helmholtz (1821-1894) and his student Wilhelm Wundt (1 832-1920).

Helmholtz applied the scientific method to the study of human vision, and his Handbook of Physiological Optics is even now described as “the single most important treatise on the physics and physiology of human vision” (Nalwa, 1993, p.15).

In 1879, Wundt opened the first laboratory of experimental psychology at the University of Leipzig. 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.

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.

Behaviorists emphasize on studying only objective measures of the percepts given to an animal and its resulting actions (or response).

The so-called mental constructs such as knowledge, beliefs, goals, and reasoning steps were dismissed as unscientific “folk psychology.”

Behaviorism discovered a lot about rats and pigeons, but had less success at understanding humans.

The view of the brain as an information-processing device, which is a principal characteristic of cognitive psychology, can be traced back at least to the works of William James.

Cambridge’s Applied Psychology Unit reestablished the legitimacy of “mental” terms as beliefs and goals, arguing that they are just as scientific as, say, using pressure and temperature to talk about gases, despite their being made of molecules that have neither.

Craik specified the three key steps of a knowledge-based agent:

1. the stimulus must be translated into an internal representation,
2. the representation is manipulated by cognitive processes to derive new internal representations, and
3. these are in turn retranslated back into action.

He clearly explained why this was a good design for an agent:

If the organism carries a “small-scale model” of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it. (Craik, 1943)

Meanwhile, in the United States, the development of computer modeling led to the creation of the field of cognitive science. The field can be said to have started at a workshop in September 1956 at MIT. This is just two months after the conference at which AI itself was “born.”

At the workshop, George Miller presented The Magic Number Seven, Noam Chomsky presented Three Models of Language, and Allen Newell and Herbert Simon presented The Logic Theory Machine.

These three influential papers showed how computer models could be used to address the psychology of memory, language, and logical thinking, respectively.

It is now a common view among psychologists that “a cognitive theory should be like a computer program” that is, it should describe a detailed information-processing mechanism whereby some cognitive function might be implemented.
Linguistics (1957-present)

How does language relate to thought?

In 1957, B. F. Skinner published *Verbal Behavior*. This was a comprehensive, detailed account of the behaviorist approach to language learning, written by the foremost expert in the field.

A review of the book became as well known as the book itself, and served to almost kill off interest in behaviorism. The author of the review was Noam Chomsky, who had just published a book on his own theory, *Syntactic Structures*.

Chomsky showed how 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 (b.c.350), 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.

The problem of understanding language soon turned out to be considerably more complex than it seemed in 1957.

*Understanding language requires an understanding of the subject matter and context, not just an understanding of the structure of sentences.*

This might seem obvious, but it was not widely appreciated until the 1960s. 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.

With the background material behind us, we are ready to cover the development of AI itself.

The State of the Art

What can AI do today?

A concise answer is difficult, because there are so many activities in so many subfields.

**Autonomous planning and scheduling**

A hundred million miles from Earth, NASA’s Remote Agent Program became the first on-board autonomous planning program. It control the scheduling of operations for a spacecraft.

Remote Agent generated plans from high-level goals specified from the ground, and it monitored the operation of the spacecraft as the plans were executed-detecting, diagnosing, and recovering from problems as they occurred.
Game playing

IBM’s Deep Blue became the first computer program to defeat the world champion in a chess match when it bested Garry Kasparov by a score of 3.5 to 2.5 in an exhibition match.

Kasparov said that he felt a “new kind of intelligence” across the board from him. *Newsweek* magazine described the match as “The brain’s last stand.” The value of IBM’s stock increased by $18 billion.

Autonomous control

The ALVINN computer vision system was trained to steer a car to keep it following a lane. Video cameras that transmit road images to ALVINN, which then computes the best direction to steer, based on experience from previous training runs.

Diagnosis

Medical diagnosis programs based on probabilistic analysis have been able to perform at the level of an expert physician in several areas of medicine. Heckerman (1991) describes a case where a leading expert on lymph-node pathology scoffs at a program’s diagnosis of an especially difficult case.

The creators of the program suggest he ask the computer for an explanation of the diagnosis. The machine points out the major factors influencing its decision and explains the subtle interaction of several of the symptoms in this case. Eventually, the expert agrees with the program.

Logistics Planning

During the Persian Gulf crisis of 1991, U.S. forces deployed a *Dynamic Analysis and Replanning Tool*, DART (Cross and Walker, 1994), to do automated logistics planning and scheduling for transportation. This involved up to 50,000 vehicles, cargo, and people at a time, and had to account for starting points, destinations, routes, and conflict resolution among all parameters.

The AI planning techniques allowed a plan to be generated in hours that would have taken weeks with older methods. The *Defense Advanced Research Project Agency* (DARPA) stated that this single application more than paid back DARPA’s 30-year investment in AI.

Robotics

Many surgeons now use robot assistants in microsurgery. Computer vision techniques are used to create a three-dimensional model of a patient’s internal anatomy and then uses robotic control to guide the insertion of a hip replacement prosthesis.

Language understanding and problem solving

PROVERB is a computer program that solves crossword puzzles better than most humans, using constraints on possible word fillers, a large database of past puzzles, and a variety of information sources including dictionaries and online databases such as a list of movies and the actors that appear in them.
Summary

Some of the important points are as follows:

Different people think of AI differently. Two important questions to ask are: Are you concerned with thinking or behavior? Do you want to model humans or work from an ideal standard?

Intelligence is concerned mainly with rational action. Ideally, an intelligent agent takes the best possible action in a situation. We will study the problem of building agents that are intelligent in this sense.

Philosophers (going back to 400 B.C.) made AI conceivable by considering the ideas that the mind is in some ways like a machine, that it operates on knowledge encoded in some internal language, and that thought can be used to choose what actions to take.

Mathematicians provided the tools to manipulate statements of logical certainty as well as uncertain, probabilistic statements. They also set the groundwork for understanding computation and reasoning about algorithms.

Economists formalized the problem of making decisions that maximize the expected outcome to the decision-maker.

Psychologists adopted the idea that humans and animals can be considered information processing machines.

Linguists showed that language use fits into this model.

Computer engineers provided the artifacts that make AI applications possible. AI programs tend to be large, and they could not work without the great advances in speed and memory that the computer industry has provided.

Control theory deals with designing devices that act optimally on the basis of feedback from the environment. Initially, the mathematical tools of control theory were quite different from AI, but the fields are coming closer together.

The history of AI has had cycles of success, misplaced optimism, and resulting cutbacks in enthusiasm and funding. There have also been cycles of introducing new creative approaches and systematically refining the best ones.

Questions

1.1 Define in your own words: (a) intelligence, (b) artificial intelligence, (c) agent.

1.2 Read Turing’s original paper on AI (Turing, 1950). In the paper, he discusses several potential objections to his proposed enterprise and his test for intelligence. Which objections still carry some weight? Are his refutations valid? Can you think of new objections arising from developments since he wrote the paper? In the paper, he predicts that, by the year 2000, a computer will have a 30% chance of passing a five-minute Turing Test with an unskilled interrogator. What chance do you think a computer would have today? In another 50 years?

1.3 There are well-known classes of problems that are intractably difficult for computers, and other classes that are provably undecidable. Does this mean that AI is impossible?

1.4 How could introspection-reporting on one’s inner thoughts be inaccurate? Could I be wrong about what I’m thinking? Discuss.
1.5 Do you think that present can or future computers could do any of the following works?
1. Playing a decent game of table tennis (ping-pong).
2. Driving in the center of Cairo.
3. Buying a week’s worth of groceries at the market.
4. Buying a week’s worth of groceries on the web.
5. Playing a decent game of bridge at a competitive level.
6. Discovering and proving new mathematical theorems.
7. Writing an intentionally funny story.
8. Giving competent legal advice in a specialized area of law.
9. Translating spoken English into spoken Swedish in real time.
10. Performing a complex surgical operation.

For the currently infeasible tasks, try to find out what the difficulties are and predict when, if ever, they will be overcome.

1.6 Some authors have claimed that perception and motor skills are the most important part of intelligence, and that “‘higher level’ capacities are necessarily parasitic-simple add-ons to these underlying facilities. Certainly, most of evolution and a large part of the brain have been devoted to perception and motor skills, whereas AI has found tasks such as game playing and logical inference to be easier, in many ways, than perceiving and acting in the real world. Do you think that AI’s traditional focus on higher-level cognitive abilities is misplaced?

1.7 Why would evolution tend to result in systems that act rationally? What goals are such systems designed to achieve?

1.8 Are reflex actions (such as moving your hand away from a hot stove) rational? Are they intelligent?

1.11 “Surely computers cannot be intelligent. They can do only what their programmers tell them.” Is the latter statement true, and does it imply the former?

1.12 “Surely animals cannot be intelligent. They can do only what their genes tell them.” Is the latter statement true, and does it imply the former?

1.13 “Surely animals, humans, and computers cannot be intelligent. They can do only what their constituent atoms are told to do by the laws of physics.” Is the latter statement true, and does it imply the former?